

# Jar Decoding: Non-Asymptotic Converse Coding Theorems, Taylor-Type Expansion, and Optimality

En-hui Yang and Jin Meng

arXiv:1204.3658v1 [cs.IT] 16 Apr 2012

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grant RGPIN203035-11, and by the Canada Research Chairs Program.

En-hui Yang and Jin Meng are with the Dept. of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. Email: [ehyang@uwaterloo.ca](mailto:ehyang@uwaterloo.ca), [j4meng@uwaterloo.ca](mailto:j4meng@uwaterloo.ca)

## Abstract

Recently, a new decoding rule called jar decoding was proposed, under which the decoder first forms a set of suitable size, called a jar, consisting of sequences from the channel input alphabet considered to be closely related to the received channel output sequence through the channel, and then takes any codeword from the jar as the estimate of the transmitted codeword; under jar decoding, a non-asymptotic achievable tradeoff between the coding rate and word error probability was also established for any discrete input memoryless channel with discrete or continuous output (DIMC). Along the path of non-asymptotic analysis, in this paper, it is further shown that jar decoding is actually optimal up to the second order coding performance by establishing new non-asymptotic converse coding theorems, and determining the (best) coding performance of finite block length for any block length  $n$  and word error probability  $\epsilon$  up to the second order. Specifically, a new converse proof technique dubbed the outer mirror image of jar is first presented and used to establish new non-asymptotic converse coding theorems for any encoding and decoding scheme. To determine the coding performance of finite block length for any block length  $n$  and error probability  $\epsilon$ , a quantity  $\delta_{t,n}(\epsilon)$  is then defined to measure the relative magnitude of the error probability  $\epsilon$  and block length  $n$  with respect to a given channel and an input distribution  $t$ . By combining the achievability of jar decoding and the new converses, it is demonstrated that when  $\epsilon < 1/2$ , the best channel coding rate  $R_n(\epsilon)$  given  $n$  and  $\epsilon$  has a “Taylor-type expansion” with respect to  $\delta_{t,n}(\epsilon)$ , where the first two terms of the expansion are  $\max_t [I(t; P) - \delta_{t,n}(\epsilon)]$ , which is equal to  $I(t^*, P) - \delta_{t^*,n}(\epsilon)$  for some optimal distribution  $t^*$ , and the third order term of the expansion is  $O(\delta_{t^*,n}^2(\epsilon))$  whenever  $\delta_{t^*,n}(\epsilon) = \Omega(\sqrt{\ln n/n})$ , thus implying the optimality of jar decoding up to the second order coding performance. Finally, based on the Taylor-type expansion and the new converses, two approximation formulas for  $R_n(\epsilon)$  (dubbed “SO” and “NEP”) are provided; they are further evaluated and compared against some of the best bounds known so far, as well as the normal approximation of  $R_n(\epsilon)$  revisited recently in the literature. It turns out that while the normal approximation is all over the map, i.e. sometime below achievable bounds and sometime above converse bounds, the SO approximation is much more reliable as it is always below converses; in the meantime, the NEP approximation is the best among the three and always provides an accurate estimation for  $R_n(\epsilon)$ . An important implication arising from the Taylor-type expansion of  $R_n(\epsilon)$  is that in the practical non-asymptotic regime, the optimal marginal codeword symbol distribution is not necessarily a capacity achieving distribution.

## Index Terms

Channel capacity, channel coding, jar decoding, non-asymptotic coding theorems, non-asymptotic equipartition properties, non-asymptotic information theory, Taylor-type expansion.

## I. INTRODUCTION

Recently, a new decoding rule called jar decoding was proposed in [1], [2], under which the decoder first forms a set of suitable size, called a jar, consisting of sequences from the channel input alphabet considered to be closely related to the received channel output sequence through the channel, and then takes any codeword from the jar as the estimate of the transmitted codeword. It was shown in [1] and [2] that under jar decoding, for any binary input memoryless channel with discrete or continuous output and with uniform capacity achieving distribution (BIMC), linear codes  $\mathcal{C}_n$  of block length  $n$  with rate  $R(\mathcal{C}_n)$  and word error probability  $P_e(\mathcal{C}_n)$  exist such that

$$P_e(\mathcal{C}_n) \leq \left( \bar{\xi}_H(X|Y, \lambda, n) + \frac{2(1 - C_{BE})M_H(X|Y, \lambda)}{\sqrt{n}\sigma_H^3(X|Y, \lambda)} \right) e^{-nr_{X|Y}(\delta)} \quad (1.1)$$

and

$$R(\mathcal{C}_n) \geq C_{\text{BIMC}} - \delta - r_{X|Y}(\delta) + \frac{\ln \frac{2(1 - C_{BE})M_H(X|Y, \lambda)}{\sqrt{n}\sigma_H^3(X|Y, \lambda)}}{n} \quad (1.2)$$

for any  $\delta \in (0, \Delta^*(X|Y))$ , where  $C_{\text{BIMC}}$  is the capacity of the given BIMC,  $\lambda = r'_{X|Y}(\delta)$ , and all other quantities are defined later in Sections II and IV. Similar achievable results were also established in [1] for non-linear codes for any discrete input memoryless channel with discrete or continuous output (DIMC).

The achievability given in (1.1) and (1.2) is quite sharp. It implies [1], [2] that for any BIMC, there exist linear codes  $\mathcal{C}_n$  of block length  $n$  such that

$$R(\mathcal{C}_n) \geq C_{\text{BIMC}} - \sigma_H(X|Y) \sqrt{\frac{2\alpha \ln n}{n}} - \left( \alpha + \frac{1}{2} \right) \frac{\ln n}{n} - O\left(\frac{\ln \ln n}{n}\right) \quad (1.3)$$

while maintaining the word error probability

$$P_e(\mathcal{C}_n) \leq \frac{n^{-\alpha}}{2\sqrt{\pi\alpha \ln n}} + O\left(n^{-\alpha} \frac{\ln n}{\sqrt{n}}\right) = \Theta\left(\frac{n^{-\alpha}}{\sqrt{\ln n}}\right) \quad (1.4)$$

and

$$R(\mathcal{C}_n) \geq C_{\text{BIMC}} - \frac{c}{\sqrt{n}} - \frac{\ln n}{2n} + \frac{1}{n} \ln \frac{(1 - C_{BE})M_H(X|Y)}{\sigma_H^3(X|Y)} \quad (1.5)$$

while maintaining the word error probability

$$P_e(\mathcal{C}_n) \leq Q\left(\frac{c}{\sigma_H(X|Y)}\right) + \frac{M_H(X|Y)}{\sigma_H^3(X|Y)} \frac{1}{\sqrt{n}}, \quad (1.6)$$

where  $\sigma_H^2(X|Y)$  and  $M_H(X|Y)$  are parameters related to the channel and specified in Section II,

$$Q(z) = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-t^2/2} dt, \quad (1.7)$$

and  $C_{BE} < 1$  is the universal constant in the Berry-Esseen central limit theorem. Furthermore, when the error probability is maintained constant in (1.6), the first two terms (i.e.,  $C_{\text{BIMC}}$  and  $\frac{c}{\sqrt{n}}$ ) in (1.5) coincide with the asymptotic second order coding rate analysis in [3], [4], [5]. Consequently, jar decoding is shown to be second order optimal asymptotically when the error probability  $\epsilon$  is maintained constant with respect to block length  $n$ .

In the non-asymptotic regime, however, the concept of constant error probability with respect to block length  $n$  is not applicable. For example, suppose that  $n = 1000$  and the error probability  $\epsilon$  is equal to  $10^{-6}$ . How would one interpret the relationship between  $\epsilon$  and  $n$  in this case? Does it make sense to interpret  $\epsilon$  as a constant with respect  $n$ ? Or is it better to interpret  $\epsilon$  as a polynomial function of  $n$ , namely,  $\epsilon = n^{-2}$ ? Since  $\epsilon$  is pretty small relatively to  $n$ , we believe that the latter interpretation makes a lot of sense in this particular case. In general, when both the error probability  $\epsilon$  and block length  $n$  are finite, what really matters is their relative magnitude to each other. Therefore, it is interesting to see if the achievability in (1.1) and (1.2) remains tight up to the second order in the non-asymptotic regime where both the error probability  $\epsilon$  and block length  $n$  are finite.

In this paper, we provide an affirmative answer to the above question. Specifically, we first present a new converse proof technique dubbed the outer mirror image of jar and use the technique to establish new non-asymptotic converse coding theorems for any binary input memoryless symmetric channel with discrete or continuous output (BIMSC) and any DIMC. We then introduce a quantity  $\delta_{t,n}(\epsilon)$  to measure the relative magnitude of the error probability  $\epsilon$  and block length  $n$  with respect to a given channel and an input distribution  $t$ . By combining the achievability of jar decoding (see (1.1) and (1.2) in the case of BIMSC) with the new converses, we further show that when  $\epsilon < 1/2$ , the best channel coding rate  $R_n(\epsilon)$  given  $n$  and  $\epsilon$  has a ‘‘Taylor-type expansion’’ with respect to  $\delta_{t,n}(\epsilon)$  in a neighborhood of  $\delta_{t,n}(\epsilon) = 0$ , where the first two terms of the expansion are  $\max_t [I(t; P) - \delta_{t,n}(\epsilon)]$ , which is equal to  $I(t^*, P) - \delta_{t^*,n}(\epsilon)$  for some optimal distribution  $t^*$ , and the third order term of the expansion is  $O(\delta_{t^*,n}^2(\epsilon))$  whenever  $\delta_{t^*,n}(\epsilon) = \Omega(\sqrt{\ln n/n})$ . Since the leading two terms in the achievability of jar decoding (see (1.2) in the case of BIMSC when  $P_e(\mathcal{C}_n) = \epsilon$ ) coincide with the first two terms of this Taylor-type expansion of  $R_n(\epsilon)$ , jar decoding is indeed optimal up to the second order coding performance in the non-asymptotical regime.

Finally, based on the Taylor-type expansion of  $R_n(\epsilon)$  and our new non-asymptotic converses,

we also derive two approximation formulas (dubbed “SO” and “NEP”) for  $R_n(\epsilon)$  in the non-asymptotic regime. The SO approximation formula consists only of the first two terms in the Taylor-type expansion of  $R_n(\epsilon)$ . On the other hand, in addition to the first two terms in the Taylor-type expansion of  $R_n(\epsilon)$ , the NEP approximation formula includes some higher order terms from our non-asymptotic converses as well. (Here, NEP stands for non-asymptotic equipartition properties established recently in [6], and underlies both the achievability bounds in (1.1) and (1.2) and our non-asymptotic converses.) These formulas are further evaluated and compared against some of the best bounds known so far, as well as the normal approximation of  $R_n(\epsilon)$  in [5]. It turns out that while the normal approximation is all over the map, i.e. sometime below achievability and sometime above converse, the SO approximation is much more reliable as it is always below converses; in the meantime, the NEP approximation is the best among the three and always provides an accurate estimation for  $R_n(\epsilon)$ . An important implication arising from the Taylor-type expansion of  $R_n(\epsilon)$  is that in the practical non-asymptotic regime, the optimal marginal codeword symbol distribution is not necessarily a capacity achieving distribution.

The rest of this paper is organized as follows. Non-asymptotic converses and the Taylor-type expansion of  $R_n(\epsilon)$  for BIMSC and DIMC are established in Sections II and III, respectively. The SO and NEP approximation formulas are developed, numerically calculated, and compared against the normal approximation in Section IV for the binary symmetric channel (BSC), binary erasure channel (BEC), binary input additive Gaussian channel (BIAGC), and Z-channel. And finally conclusions are drawn in Section V.

## II. NON-ASYMPTOTIC CONVERSE AND TAYLOR-TYPE EXPANSION: BIMSC

Consider a BIMC  $\{p(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ , where  $\mathcal{X} = \{0, 1\}$  is the channel input alphabet, and  $\mathcal{Y}$  is the channel output alphabet, which is arbitrary and could be discrete or continuous. Throughout this section, let  $X$  denote the uniform random variable on  $\mathcal{X}$  and  $Y$  the corresponding channel output of the BIMC in response to  $X$ . Then the capacity (in nats) of the BIMC is calculated by

$$C_{\text{BIMC}} = \ln 2 - H(X|Y) \quad (2.1)$$

where  $H(X|Y)$  is the conditional entropy of  $X$  given  $Y$ . Here and throughout the rest of the paper,  $\ln$  stands for the logarithm with base  $e$ , and all information quantities are measured in

nats. Further assume that the random variable  $-\ln p(0|Y)$  given  $X = 0$  and the random variable  $-\ln p(1|Y)$  given  $X = 1$  have the same distribution, where  $p(0|Y)$  ( $p(1|Y)$ , respectively) denotes the conditional probability of  $X = 0$  ( $X = 1$ , respectively) given  $Y$ . Such a BIMC is called a binary input memoryless symmetrical channel (BIMSC). (It can be verified that BSC, BEC, BIAGC, and general binary input symmetric output channels all belong to the class of BIMSC.) Under this assumption, we have

$$\Pr \left\{ -\frac{1}{n} \ln p(X^n|Y^n) > H(X|Y) + \delta \middle| X^n = x^n \right\} = \Pr \left\{ -\frac{1}{n} \ln p(X^n|Y^n) > H(X|Y) + \delta \right\} \quad (2.2)$$

for any  $x^n \in \mathcal{X}^n$ , where  $Y^n$  is the output of the BIMSC in response to  $X^n$ , the  $n$  independent copies of  $X$ . Throughout this paper, for any set  $S$ , we use  $S^n$  to denote the set of all sequences of length  $n$  drawn from  $S$ .

#### A. Definitions

Before stating our converse channel coding theorem for the BIMSC, let us first introduce some definitions from [6]. Define

$$\lambda^*(X|Y) \triangleq \sup \left\{ \lambda \geq 0 : \int p(y) \left[ \sum_{x \in \mathcal{X}} p^{-\lambda+1}(x|y) \right] dy < \infty \right\} \quad (2.3)$$

where  $\int dy$  is understood throughout this paper to be the summation over  $\mathcal{Y}$  if  $\mathcal{Y}$  is discrete. Suppose that

$$\lambda^*(X|Y) > 0. \quad (2.4)$$

Define for any  $\delta \geq 0$

$$r_{X|Y}(\delta) \triangleq \sup_{\lambda \geq 0} \left[ \lambda(H(X|Y) + \delta) - \ln \sum_{x \in \mathcal{X}} \int p(y) p^{-\lambda+1}(x|y) dy \right]. \quad (2.5)$$

For any  $\lambda \in [0, \lambda^*(X|Y))$ , let  $X_\lambda$  and  $Y_\lambda$  be random variables under joint distribution  $p(x, y) f_\lambda(x, y)$  where

$$f_\lambda(x, y) \triangleq \frac{p^{-\lambda}(x|y)}{\sum_{u \in \mathcal{X}} \int p(v) p^{-\lambda+1}(u|v) dv}. \quad (2.6)$$

Further define

$$\delta(\lambda) \triangleq \mathbf{E}[-\ln p(X_\lambda|Y_\lambda)] - H(X|Y) \quad (2.7)$$

$$\Delta^*(X|Y) \triangleq \lim_{\lambda \uparrow \lambda^*(X|Y)} \delta(\lambda) \quad (2.8)$$

$$\sigma_H^2(X|Y, \lambda) \triangleq \mathbf{Var}[-\ln p(X_\lambda|Y_\lambda)] = \mathbf{E}[|-\ln p(X_\lambda|Y_\lambda) - \mathbf{E}[-\ln p(X_\lambda|Y_\lambda)]|^2] \quad (2.9)$$

$$M_H(X|Y, \lambda) \triangleq \mathbf{M}_3[-\ln p(X_\lambda|Y_\lambda)] = \mathbf{E}[|-\ln p(X_\lambda|Y_\lambda) - \mathbf{E}[-\ln p(X_\lambda|Y_\lambda)]|^3] \quad (2.10)$$

and

$$\hat{M}_H(X|Y, \lambda) \triangleq \hat{\mathbf{M}}_3[-\ln p(X_\lambda|Y_\lambda)] = \mathbf{E}[-\ln p(X_\lambda|Y_\lambda) - \mathbf{E}[-\ln p(X_\lambda|Y_\lambda)]]^3 \quad (2.11)$$

where  $\mathbf{E}[\cdot]$ ,  $\mathbf{Var}[\cdot]$ ,  $\mathbf{M}_3[\cdot]$ , and  $\hat{\mathbf{M}}_3[\cdot]$  are respectively expectation, variance, third absolute central moment, and third central moment operators on random variables, and write  $\hat{M}_H(X|Y, 0)$  as  $\hat{M}_H(X|Y)$ ,  $M_H(X|Y, 0)$  as  $M_H(X|Y)$ , and  $\sigma_H^2(X|Y, 0)$  as  $\sigma_H^2(X|Y)$ . Clearly,  $\sigma_H^2(X|Y)$ ,  $M_H(X|Y)$ , and  $\hat{M}_H(X|Y)$  are the variance, third absolute central moment, and third central moment of  $-\ln p(X|Y)$ . In particular,  $\sigma_H^2(X|Y)$  is referred to as the conditional information variance of  $X$  given  $Y$  in [6]. Assume that

$$\sigma_H^2(X|Y) > 0 \text{ and } M_H(X|Y) = \mathbf{M}_3[-\ln p(X|Y)] < \infty. \quad (2.12)$$

Then it follows from [6] that  $r_{X|Y}(\delta)$  is strictly increasing, convex, and continuously differentiable up to at least the third order inclusive over  $\delta \in [0, \Delta^*(X|Y))$ , and furthermore has the following parametric expression

$$r_{X|Y}(\delta(\lambda)) = \lambda(H(X|Y) + \delta(\lambda)) - \ln \sum_{x \in \mathcal{X}} \int p(y) p^{-\lambda+1}(x|y) dy \quad (2.13)$$

with  $\delta(\lambda)$  defined in (2.7) and  $\lambda = r'_{X|Y}(\delta)$ . In addition, let

$$\begin{aligned} \bar{\xi}_H(X|Y, \lambda, n) &\triangleq \frac{2C_{BE}M_H(X|Y, \lambda)}{\sqrt{n}\sigma_H^3(X|Y, \lambda)} \\ &+ e^{\frac{n\lambda^2\sigma_H^2(X|Y, \lambda)}{2}} [Q(\sqrt{n}\lambda\sigma_H(X|Y, \lambda)) - Q(\rho^* + \sqrt{n}\lambda\sigma_H(X|Y, \lambda))] \end{aligned} \quad (2.14)$$

$$\underline{\xi}_H(X|Y, \lambda, n) \triangleq e^{\frac{n\lambda^2\sigma_H^2(X|Y, \lambda)}{2}} Q(\rho_* + \sqrt{n}\lambda\sigma_H(X|Y, \lambda)) \quad (2.15)$$

with  $Q(\rho^*) = \frac{C_{BE}M_H(X|Y, \lambda)}{\sqrt{n}\sigma_H^3(X|Y, \lambda)}$  and  $Q(\rho_*) = \frac{1}{2} - \frac{2C_{BE}M_H(X|Y, \lambda)}{\sqrt{n}\sigma_H^3(X|Y, \lambda)}$ .

The significance of the above quantities related to the channel can be seen from Theorem 4 in [6], summarized as below:

(a) *There exists a  $\delta^* > 0$  such that for any  $\delta \in (0, \delta^*]$ ,*

$$r_{X|Y}(\delta) = \frac{1}{2\sigma_H^2(X|Y)}\delta^2 + O(\delta^3). \quad (2.16)$$

(b) For any  $\delta \in (0, \Delta^*(X|Y))$  and any positive integer  $n$

$$\begin{aligned}\bar{\xi}_H(X|Y, \lambda, n)e^{-nr_{X|Y}(\delta)} &\geq \Pr \left\{ -\frac{1}{n} \ln p(X^n|Y^n) > H(X|Y) + \delta \right\} \\ &\geq \underline{\xi}_H(X|Y, \lambda, n)e^{-nr_{X|Y}(\delta)},\end{aligned}\quad (2.17)$$

where  $\lambda = r'_{X|Y}(\delta) > 0$ . Moreover, when  $\delta = o(1)$  and  $\delta = \Omega(1/\sqrt{n})$ ,

$$\bar{\xi}_H(X|Y, \lambda, n) = e^{\frac{n\lambda^2\sigma_H^2(X|Y, \lambda)}{2}} Q(\sqrt{n}\lambda\sigma_H(X|Y, \lambda)) (1 + o(1)) \quad (2.18)$$

$$\underline{\xi}_H(X|Y, \lambda, n) = e^{\frac{n\lambda^2\sigma_H^2(X|Y, \lambda)}{2}} Q(\sqrt{n}\lambda\sigma_H(X|Y, \lambda)) (1 - o(1)) \quad (2.19)$$

and

$$e^{\frac{n\lambda^2\sigma_H^2(X|Y, \lambda)}{2}} Q(\sqrt{n}\lambda\sigma_H(X|Y, \lambda)) = \Theta\left(\frac{1}{\sqrt{n}\lambda}\right) \quad (2.20)$$

with  $\lambda = r'_X(\delta) = \Theta(\delta)$ .

(c) For any  $\delta \leq c\sqrt{\frac{\ln n}{n}}$ , where  $c < \sigma_H(X|Y)$  is a constant,

$$\begin{aligned}Q\left(\frac{\delta\sqrt{n}}{\sigma_H(X|Y)}\right) - \frac{C_{BE}M_H(X|Y)}{\sqrt{n}\sigma_H^3(X|Y)} &\leq \Pr \left\{ -\frac{1}{n} \ln p(X^n|Y^n) > H(X|Y) + \delta \right\} \\ &\leq Q\left(\frac{\delta\sqrt{n}}{\sigma_H(X|Y)}\right) + \frac{C_{BE}M_H(X|Y)}{\sqrt{n}\sigma_H^3(X|Y)}.\end{aligned}\quad (2.21)$$

Define for any  $x^n \in \mathcal{X}^n$ ,

$$B(x^n, \delta) \triangleq \left\{ y^n : \infty > -\frac{1}{n} \ln p(x^n|y^n) > H(X|Y) + \delta \right\} \quad (2.22)$$

and

$$B_{n,\delta} \triangleq \cup_{x^n \in \mathcal{X}^n} B(x^n, \delta). \quad (2.23)$$

Since for any  $y^n \in \mathcal{Y}^n$ , the following set

$$\left\{ x^n \in \mathcal{X}^n : -\frac{1}{n} \ln p(x^n|y^n) \leq H(X|Y) + \delta \right\} \quad (2.24)$$

is referred to as a BIMC jar for  $y^n$  in [1], [2], we shall call  $B(x^n, \delta)$  the *outer mirror image of jar* corresponding to  $x^n$ . Moreover, define for any set  $B \subseteq \mathcal{Y}^n$ ,

$$P(B) \triangleq \Pr \{Y^n \in B\} \quad (2.25)$$

$$P_{x^n}(B) \triangleq \Pr \{Y^n \in B | X^n = x^n\}. \quad (2.26)$$



It is easy to see that

$$\begin{aligned} P_{x^n}(B(x^n, \delta)) &= \Pr \left\{ -\frac{1}{n} \ln p(X^n|Y^n) > H(X|Y) + \delta \mid X^n = x^n \right\} \\ &= \Pr \left\{ -\frac{1}{n} \ln p(X^n|Y^n) > H(X|Y) + \delta \right\} \end{aligned} \quad (2.27)$$

where the last equality is due to (2.2).

### B. Converse Coding Theorem

We are now ready to state our non-asymptotic converse coding theorem for BIMSCs.

**Theorem 1.** *Given a BIMSC, for any channel code  $\mathcal{C}_n$  of block length  $n$  with average word error probability  $P_e(\mathcal{C}_n) = \epsilon_n$ ,*

$$R(\mathcal{C}_n) \leq C_{\text{BIMSC}} - \delta - \frac{\ln \epsilon_n - \ln P(B_{n,\delta}) + \ln \frac{-2 \ln \epsilon_n}{\sigma_H^2(X|Y)n} - \ln \left( 1 + \frac{\sqrt{\frac{-2 \ln \epsilon_n}{n}}}{\sigma_H(X|Y)} \right)}{n} \quad (2.28)$$

where  $\delta$  is the largest number such that

$$\left( 1 + \frac{2}{\sigma_H(X|Y)} \sqrt{\frac{-2 \ln \epsilon_n}{n}} \right) \epsilon_n \leq \Pr \left\{ -\frac{1}{n} \ln p(X^n|Y^n) > H(X|Y) + \delta \right\}. \quad (2.29)$$

Moreover, the following hold:

1)

$$R(\mathcal{C}_n) \leq C_{\text{BIMSC}} - \delta - \frac{\ln \epsilon_n - \ln P(B_{n,\delta}) + \ln \frac{-2 \ln \epsilon_n}{\sigma_H^2(X|Y)n} - \ln \left( 1 + \frac{\sqrt{\frac{-2 \ln \epsilon_n}{n}}}{\sigma_H(X|Y)} \right)}{n} \quad (2.30)$$

where  $\delta$  is the solution to

$$\left( 1 + \frac{2}{\sigma_H(X|Y)} \sqrt{\frac{-2 \ln \epsilon_n}{n}} \right) \epsilon_n = \xi_H(X|Y, \lambda, n) e^{-nr_{X|Y}(\delta)} \quad (2.31)$$

with  $\delta(\lambda) = \delta$ .

2) When  $\epsilon_n = \frac{e^{-n^\alpha}}{2\sqrt{\pi n^\alpha}} \left( 1 - \frac{1}{2n^\alpha} \right)$  for  $\alpha \in (0, 1)$ ,

$$R(\mathcal{C}_n) \leq C_{\text{BIMSC}} - \sqrt{2} \sigma_H(X|Y) n^{-\frac{1-\alpha}{2}} + O(n^{-(1-\alpha)}). \quad (2.32)$$

3) When  $\epsilon_n = \frac{n^{-\alpha}}{2\sqrt{\pi \alpha \ln n}} \left( 1 - \frac{1}{2\alpha \ln n} \right)$  for  $\alpha > 0$ ,

$$R(\mathcal{C}_n) \leq C_{\text{BIMSC}} - \sigma_H(X|Y) \sqrt{\frac{2\alpha \ln n}{n}} + O\left(\frac{\ln n}{n}\right). \quad (2.33)$$

4) When  $\epsilon_n = \epsilon$  satisfying  $\epsilon + \frac{1}{\sqrt{n}} \left( \frac{2\sqrt{-2\ln \epsilon}}{\sigma_H(X|Y)} \epsilon + \frac{C_{BE} M_H(X|Y)}{\sigma_H^3(X|Y)} \right) < 1$ ,

$$R(\mathcal{C}_n) \leq C_{\text{BIMSC}} - \frac{\ln \epsilon + \ln \frac{-2\ln \epsilon}{\sigma_H^2(X|Y)n} - \ln \left( 1 + \frac{\sqrt{-2\ln \epsilon}}{\sigma_H(X|Y)} \right)}{n} - \frac{\sigma_H(X|Y)}{\sqrt{n}} Q^{-1} \left( \epsilon + \frac{1}{\sqrt{n}} \left( \frac{2\sqrt{-2\ln \epsilon}}{\sigma_H(X|Y)} \epsilon + \frac{C_{BE} M_H(X|Y)}{\sigma_H^3(X|Y)} \right) \right) \quad (2.34)$$

$$= C_{\text{BIMSC}} - \frac{\sigma_H(X|Y)}{\sqrt{n}} Q^{-1}(\epsilon) + \frac{\ln n}{n} + O(n^{-1}). \quad (2.35)$$

*Proof:* Assume that the message  $M$  is uniformly distributed in  $\{1, 2, \dots, e^{nR(\mathcal{C}_n)}\}$ ,  $x^n(m)$  is the codeword corresponding to the message  $m$ , and  $\epsilon_{m,n}$  is the conditional error probability given message  $m$ . Then

$$\epsilon_n = \mathbf{E}[\epsilon_{M,n}]. \quad (2.36)$$

Let

$$\mathcal{M} \triangleq \{m : \epsilon_{m,n} \leq \epsilon_n(1 + \beta_n)\}, \quad (2.37)$$

where  $\beta_n > 0$  will be specified later. By Markov inequality,

$$\Pr\{M \in \mathcal{M}\} \geq \frac{\beta_n}{1 + \beta_n} \text{ and } |\mathcal{M}| \geq e^{nR(\mathcal{C}_n) + \ln \frac{\beta_n}{1 + \beta_n}}. \quad (2.38)$$

Denote the decision region for message  $m \in \mathcal{M}$  as  $D_m$ . Then

$$\begin{aligned} P_{x^n(m)}(B(x^n(m), \delta) \cap D_m) &= P_{x^n(m)}(B(x^n(m), \delta)) - P(B(x^n(m), \delta) \cap D_m^c) \\ &\geq P_{x^n(m)}(B(x^n(m), \delta)) - \epsilon_{m,n} \\ &\geq P_{x^n(m)}(B(x^n(m), \delta)) - \epsilon_n(1 + \beta_n) \\ &= \Pr \left\{ -\frac{1}{n} \ln p(X^n|Y^n) > H(X|Y) + \delta \right\} - \epsilon_n(1 + \beta_n) \end{aligned} \quad (2.39)$$

where the last equality is due to (2.27). At this point, we select  $\delta$  such that

$$\Pr \left\{ -\frac{1}{n} \ln p(X^n|Y^n) > H(X|Y) + \delta \right\} \geq \epsilon_n(1 + 2\beta_n). \quad (2.40)$$

Substituting (2.40) into (2.39), we have

$$P_{x^n(m)}(B(x^n(m), \delta) \cap D_m) \geq \beta_n \epsilon_n. \quad (2.41)$$

By the fact that  $D_m$  are disjoint for different  $m$  and

$$\cup_{m \in \mathcal{M}} (B(x^n(m), \delta) \cap D_m) \subseteq B_{n,\delta}, \quad (2.42)$$

we have

$$\begin{aligned} P(B_{n,\delta}) &= \int_{B_{n,\delta}} p(y^n) dy^n \\ &\geq \sum_{m \in \mathcal{M}} \int_{B(x^n(m), \delta) \cap D_m} p(y^n) dy^n \\ &= \sum_{m \in \mathcal{M}} \int_{B(x^n(m), \delta) \cap D_m} \frac{p(y^n | x^n(m)) p(x^n(m))}{p(x^n(m) | y^n)} dy^n \\ &\stackrel{1)}{\geq} \sum_{m \in \mathcal{M}} \int_{B(x^n(m), \delta) \cap D_m} p(y^n | x^n(m)) e^{n(-C_{\text{BIMSC}} + \delta)} dy^n \\ &= \sum_{m \in \mathcal{M}} e^{n(-C_{\text{BIMSC}} + \delta)} \int_{B(x^n(m), \delta) \cap D_m} p(y^n | x^n(m)) dy^n \\ &= \sum_{m \in \mathcal{M}} e^{n(-C_{\text{BIMSC}} + \delta)} P_{x^n(m)}(B(x^n(m), \delta) \cap D_m) \\ &\stackrel{2)}{\geq} \sum_{m \in \mathcal{M}} e^{n(-C_{\text{BIMSC}} + \delta)} \beta_n \epsilon_n = |\mathcal{M}| e^{n(-C_{\text{BIMSC}} + \delta)} \beta_n \epsilon_n \end{aligned} \quad (2.43)$$

where the inequality 1) is due to the definition of  $B(x^n, \delta)$  given in (2.22), and the inequality 2) follows from (2.41). From (2.43), it follows that

$$|\mathcal{M}| \leq e^{n(C_{\text{BIMSC}} - \delta) - \ln \beta_n - \ln \epsilon_n + \ln P(B_{n,\delta})}. \quad (2.44)$$

Then combining (2.38) and (2.44) yields

$$R(\mathcal{C}_n) \leq C_{\text{BIMSC}} - \delta - \frac{\ln \frac{\beta_n}{1+\beta_n}}{n} - \frac{\ln \beta_n}{n} - \frac{\ln \epsilon_n - \ln P(B_{n,\delta})}{n} \quad (2.45)$$

By letting  $\beta_n = \frac{1}{\sigma_H(X|Y)} \sqrt{\frac{-2 \ln \epsilon_n}{n}}$ , (2.28) and (2.29) directly come from (2.40) and (2.45).

- 1) By (2.17) shown in [6], selecting  $\delta$  to be the solution to (2.31) will make (2.40) satisfied, and therefore (2.30) is proved.
- 2) Towards proving (2.32), we want to show that by making  $\delta = \sqrt{2} \sigma_H(X|Y) n^{-\frac{1-\alpha}{2}} - \eta n^{-(1-\alpha)}$  for some constant  $\eta$ ,

$$\Pr \left\{ -\frac{1}{n} p(X^n | Y^n) > H(X|Y) + \delta \right\} \geq \left( 1 + \frac{2}{\sigma_H(X|Y)} \sqrt{\frac{-2 \ln \epsilon_n}{n}} \right) \epsilon_n \quad (2.46)$$

with  $\epsilon_n = \frac{e^{-n^\alpha}}{2\sqrt{\pi n^\alpha}} \left(1 - \frac{1}{2n^\alpha}\right)$ . Then the proof follows essentially the same approach as that of (2.33), shown below in details.

- 3) Apply the trivial bound  $P(B_{n,\delta}) \leq 1$ . Then to show (2.33), we only have to show that  $\delta = \sigma_H(X|Y)\sqrt{\frac{2\alpha \ln n}{n}} - \frac{\eta \ln n}{n}$  for some constant  $\eta$  can make

$$\begin{aligned} & \Pr \left\{ -\frac{1}{n} p(X^n|Y^n) > H(X|Y) + \delta \right\} \\ & \geq \xi_H(X|Y, \lambda, n) e^{-nr_{X|Y}(\delta)} \\ & \geq \left( 1 + \eta_0 \sqrt{\frac{\ln n}{n}} \right) \frac{n^{-\alpha}}{2\sqrt{\pi \alpha \ln n}} \left( 1 - \frac{1}{2\alpha \ln n} \right) \\ & \geq \left( 1 + \frac{2}{\sigma_H(X|Y)} \sqrt{\frac{-2 \ln \epsilon_n}{n}} \right) \epsilon_n \end{aligned} \quad (2.47)$$

satisfied, where  $\lambda = r'_{X|Y}(\delta)$  and

$$\frac{2}{\sigma_H(X|Y, \lambda)} \sqrt{\frac{-2 \ln \epsilon_n}{n}} = \Theta \left( \sqrt{\frac{\ln n}{n}} \right) \leq \eta_0 \sqrt{\frac{\ln n}{n}} \quad (2.48)$$

for some constant  $\eta_0$ . Towards this, recall (2.16) (2.19) and (2.20),

$$\begin{aligned} e^{-nr_{X|Y}(\delta)} &= e^{-nr_{X|Y}(\sigma_H(X|Y)\sqrt{\frac{2\alpha \ln n}{n}} - \frac{\eta \ln n}{n})} \\ &= e^{-n \left[ \frac{1}{2\sigma_H^2(X|Y)} \left( \sigma_H(X|Y)\sqrt{\frac{2\alpha \ln n}{n}} - \frac{\eta \ln n}{n} \right)^2 + O\left(\sqrt{\frac{\ln^3 n}{n^3}}\right) \right]} \\ &= e^{-\alpha \ln n + \frac{\eta}{\sigma_H(X|Y)} \sqrt{\frac{2\alpha \ln^3 n}{n}} - O\left(\sqrt{\frac{\ln^3 n}{n}}\right)} \\ &\geq e^{-\alpha \ln n + \left( \frac{\sqrt{2\alpha}\eta}{\sigma_H(X|Y)} - \eta_1 \right) \sqrt{\frac{\ln^3 n}{n}}} \end{aligned} \quad (2.49)$$

for some constant  $\eta_1$ , and

$$\begin{aligned} & \xi_H(X|Y, \lambda, n) \\ &= e^{\frac{n\lambda^2\sigma_H^2(X|Y,\lambda)}{2}} Q(\rho_* + \sqrt{n}\lambda\sigma_H(X|Y, \lambda)) \\ &\geq e^{\frac{n\lambda^2\sigma_H^2(X|Y,\lambda)}{2}} \frac{e^{-\frac{(\rho_* + \sqrt{n}\lambda\sigma_H(X|Y,\lambda))^2}{2}}}{\sqrt{2\pi}(\rho_* + \sqrt{n}\lambda\sigma_H(X|Y, \lambda))} \left[ 1 - \frac{1}{(\rho_* + \sqrt{n}\lambda\sigma_H(X|Y, \lambda))^2} \right] \\ &= \frac{e^{-\frac{\rho_*^2 + 2\rho_*\sqrt{n}\lambda\sigma_H(X|Y,\lambda)}{2}}}{\sqrt{2\pi}(\rho_* + \sqrt{n}\lambda\sigma_H(X|Y, \lambda))} \left[ 1 - \frac{1}{(\rho_* + \sqrt{n}\lambda\sigma_H(X|Y, \lambda))^2} \right] \\ &\geq \frac{1}{2\sqrt{\pi\alpha \ln n}} \left( 1 - \frac{1}{2\alpha \ln n} \right) \left( 1 - \Theta \left( \sqrt{\frac{\ln n}{n}} \right) \right) \\ &\geq \frac{1}{2\sqrt{\pi\alpha \ln n}} \left( 1 - \frac{1}{2\alpha \ln n} \right) \left( 1 - \eta_2 \sqrt{\frac{\ln n}{n}} \right) \end{aligned} \quad (2.50)$$

for another constant  $\eta_2$ , where  $\rho_* = Q^{-1} \left( \frac{1}{2} - \frac{2C_{BE}M_H(X|Y,\lambda)}{\sqrt{n}\sigma_H^3(X|Y,\lambda)} \right) = \Theta \left( \frac{1}{\sqrt{n}} \right)$ , and we utilize the fact that

$$\begin{aligned} \lambda &= r'_{X|Y}(\delta) \\ &= \frac{\delta}{\sigma_H^2(X|Y)} + O(\delta^2) \end{aligned} \quad (2.51)$$

$$\sigma_H(X|Y, \lambda) = \sigma_H(X|Y) \pm O(\lambda). \quad (2.52)$$

Then (2.47) is satisfied by choosing a constant  $\eta$  such that

$$\begin{aligned} &e^{\left(\frac{\sqrt{2\alpha}\eta}{\sigma_H(X|Y)} - \eta_1\right)\sqrt{\frac{\ln^3 n}{n}}} \left(1 - \eta_2\sqrt{\frac{\ln n}{n}}\right) \\ &\geq \left[1 + \left(\frac{\sqrt{2\alpha}\eta}{\sigma_H(X|Y)} - \eta_1\right)\sqrt{\frac{\ln^3 n}{n}}\right] \left(1 - \eta_2\sqrt{\frac{\ln n}{n}}\right) \\ &\geq 1 + \eta_0\sqrt{\frac{\ln n}{n}} \end{aligned} \quad (2.53)$$

for some constants  $\eta_0$ ,  $\eta_1$  and  $\eta_2$ .

4) According to (2.40), we should select  $\delta$  such that

$$\Pr \left\{ -\frac{1}{n} \ln p(X^n|Y^n) > H(X|Y) + \delta \right\} \geq \left(1 + \frac{2}{\sigma_H(X|Y)} \sqrt{\frac{-2 \ln \epsilon}{n}}\right) \epsilon. \quad (2.54)$$

Then by (2.21),

$$\delta = \frac{\sigma_H(X|Y)}{\sqrt{n}} Q^{-1} \left( \epsilon + \frac{1}{\sqrt{n}} \left( \frac{2\sqrt{-2 \ln \epsilon}}{\sigma_H(X|Y)} \epsilon + \frac{M_H(X|Y)}{\sigma_H^3(X|Y)} \right) \right) \quad (2.55)$$

will guarantee (2.54). Consequently, (2.34) is proved by substituting (2.55) and  $\epsilon_n = \epsilon$  into (2.45) and applying the trivial bound  $P(B_{n,\delta}) \leq 1$ , and (2.35) follows the fact that

$$Q^{-1} \left( \epsilon + \frac{1}{\sqrt{n}} \left( \frac{2\sqrt{-2 \ln \epsilon}}{\sigma_H(X|Y)} \epsilon + \frac{C_{BE}M_H(X|Y)}{\sigma_H^3(X|Y)} \right) \right) = Q^{-1}(\epsilon) - O \left( \frac{1}{\sqrt{n}} \right). \quad (2.56)$$

■

**Remark 1.** *It is clear that the above converse proof technique depends heavily on the concept of the outer mirror image of jar corresponding to codewords. To facilitate its future reference, it is beneficial to loosely call such a converse proof technique the outer mirror image of jar.*

**Remark 2.** In general, the evaluation of  $P(B_{n,\delta})$  may not be feasible, in which case the trivial bound  $P(B_{n,\delta}) \leq 1$  can be applied without affecting the second order performance in the non-exponential error probability regime, as shown above. However, there are cases where  $P(B_{n,\delta})$  can be tightly bounded (e.g. BEC, shown in section IV).

**Remark 3.** For the bound (2.34), when  $\epsilon$  is small with respect to  $\frac{1}{\sqrt{n}}$ ,  $\frac{C_{BE}M_H(X|Y)}{\sqrt{n}\sigma_H^3(X|Y)}$  (the estimation error that comes from Berry-Esseen central limit theorem) will be dominant; in this case, (2.34) is loose.

**Remark 4.** The choice  $\beta_n = \frac{1}{\sigma_H(X|Y)} \sqrt{\frac{-2 \ln \epsilon_n}{n}}$  in the proof of Theorem 1 is not arbitrary. Actually, it is optimal when  $\delta$  is small in the sense of minimizing the upper bound (2.45) in which  $\delta$  depends on  $\beta_n$  through (2.40). To derive the expression for  $\beta_n$ , the following approximations can be adopted when  $\delta$  is small:

$$\frac{d\delta}{d\beta_n} \approx -\frac{2\beta_n\sigma_H^2(X|Y)}{n\delta} \quad (2.57)$$

$$\delta^2 \approx \frac{-2\sigma_H^2(X|Y) \ln \epsilon_n}{n} \quad (2.58)$$

$$\ln \frac{\beta_n}{1 + \beta_n} \approx \ln \beta_n \quad (2.59)$$

where (2.57) and (2.58) can be developed from (2.16) and (2.17).

By reviewing the proof of Theorem 1, it is not hard to reach the following corollary.

**Corollary 1.** Given a BIMSC, for any channel code  $\mathcal{C}_n$  of block length  $n$  with maximum error probability  $P_m(\mathcal{C}_n) = \epsilon_n$ ,

$$R(\mathcal{C}_n) \leq C_{\text{BIMSC}} - \delta - \frac{\ln \epsilon_n + \ln \frac{1}{\sigma_H(X|Y)} \sqrt{\frac{-2 \ln \epsilon_n}{n}} - \ln P(B_{n,\delta})}{n} \quad (2.60)$$

where  $\delta$  is the largest number such that

$$\left(1 + \frac{1}{\sigma_H(X|Y)} \sqrt{\frac{-2 \ln \epsilon_n}{n}}\right) \epsilon_n \leq \Pr \left\{ -\frac{1}{n} \ln p(X^n|Y^n) > H(X|Y) + \delta \right\}. \quad (2.61)$$

Moreover, the following hold:

1)

$$R(\mathcal{C}_n) \leq C_{\text{BIMSC}} - \delta - \frac{\ln \epsilon_n + \ln \frac{1}{\sigma_H(X|Y)} \sqrt{\frac{-2 \ln \epsilon_n}{n}} - \ln P(B_{n,\delta})}{n} \quad (2.62)$$

where  $\delta$  is the solution to

$$\left(1 + \frac{1}{\sigma_H(X|Y)} \sqrt{\frac{-2 \ln \epsilon_n}{n}}\right) \epsilon_n = \underline{\xi}_H(X|Y, \lambda, n) e^{-nr_{X|Y}(\delta)} \quad (2.63)$$

with  $\delta(\lambda) = \delta$ .

- 2) When  $\epsilon_n = \epsilon$  satisfying  $\epsilon + \frac{1}{\sqrt{n}} \left( \frac{\sqrt{-2 \ln \epsilon}}{\sigma_H(X|Y)} \epsilon + \frac{C_{BE} M_H(X|Y)}{\sigma_H^3(X|Y)} \right) < 1$ ,

$$\begin{aligned} R(\mathcal{C}_n) &\leq C_{\text{BIMSC}} - \frac{\ln \epsilon + \ln \frac{1}{\sigma_H(X|Y)} \sqrt{\frac{-2 \ln \epsilon}{n}}}{n} \\ &\quad - \frac{\sigma_H(X|Y)}{\sqrt{n}} Q^{-1} \left( \epsilon + \frac{1}{\sqrt{n}} \left( \frac{\sqrt{-2 \ln \epsilon}}{\sigma_H(X|Y)} \epsilon + \frac{C_{BE} M_H(X|Y)}{\sigma_H^3(X|Y)} \right) \right) \end{aligned} \quad (2.64)$$

$$= C_{\text{BIMSC}} - \frac{\sigma_H(X|Y)}{\sqrt{n}} Q^{-1}(\epsilon) + \frac{\ln n}{2n} + O(n^{-1}) \quad (2.65)$$

Remarks 2, 3 and 4 also apply to Corollary 1.

### C. Taylor-type Expansion

Fix a BIMSC. For any block length  $n$  and average error probability  $\epsilon$ , let  $R_n(\epsilon)$  be the best coding rate achievable with block length  $n$  and average error probability  $\leq \epsilon$ , i.e.,

$$R_n(\epsilon) \triangleq \max\{R(\mathcal{C}_n) : \mathcal{C}_n \text{ is a channel code of block length } n \text{ with } P_e(\mathcal{C}_n) \leq \epsilon\}. \quad (2.66)$$

In this subsection, we combine the non-asymptotic achievability given in (1.1) (1.2) with the non-asymptotic converses given in (2.28) to (2.31) to derive a Taylor-type expansion of  $R_n(\epsilon)$  in the non-asymptotic regime where both  $n$  and  $\epsilon$  are finite. As mentioned early, when both  $n$  and  $\epsilon$  are finite, what really matters is the relative magnitude of  $\epsilon$  and  $n$ . As such, we begin with introducing a quantity  $\delta_n(\epsilon)$  to measure the relative magnitude of  $\epsilon$  and  $n$  with respect to the given BIMSC.

A close look at the non-asymptotic achievability given in (1.1) (1.2) and the non-asymptotic converses given in (2.28) to (2.31) reveals that

$$\Pr \left\{ -\frac{1}{n} \ln p(X^n|Y^n) > H(X|Y) + \delta \right\}$$

is crucial in both cases. According to (2.18) and (2.19),

$$\begin{aligned} \Pr \left\{ -\frac{1}{n} \ln p(X^n|Y^n) > H(X|Y) + \delta \right\} &\approx e^{\frac{n\lambda^2 \sigma_H^2(X|Y, \lambda)}{2}} Q(\sqrt{n} \lambda \sigma_H(X|Y, \lambda)) e^{-nr_{X|Y}(\delta)} \\ &\triangleq g_{X|Y, n}(\delta) \end{aligned} \quad (2.67)$$

where  $\lambda = r'_{X|Y}(\delta)$ . Consequently, we would like to define  $\delta_n(\epsilon)$  as the solution to

$$g_{X|Y,n}(\delta) = \epsilon \quad (2.68)$$

given  $n$  and  $\epsilon \leq 1/2$ , where the uniqueness of the solution in certain range is shown in Lemma 1.

**Lemma 1.** *There exists  $\delta^+ > 0$  such that for any  $n > 0$ ,  $g_{X|Y,n}(\delta)$  is a strictly decreasing function of  $\delta$  over  $\delta \in [0, \delta^+]$ .*

*Proof:* Since  $\lambda = r'_{X|Y}(\delta)$ , it follows from (2.7) and (2.13) that  $g_{X|Y,n}(\delta) = g_{X|Y,n}(\delta(\lambda))$  is a function of  $\lambda$  through  $\delta = \delta(\lambda)$ . (For details about the properties of  $\delta(\lambda)$  and  $r_{X|Y}(\delta)$ , please see [6].) Moreover, by the fact that  $\delta(0) = 0$  and  $\delta(\lambda)$  is a strictly increasing function of  $\lambda$ , the proof of this lemma is yielded by analyzing the derivative of  $g_{X|Y,n}(\delta(\lambda))$  with respect to  $\lambda$  around  $\lambda = 0$ . Towards this,

$$\begin{aligned} & \frac{dg_{X|Y,n}(\delta(\lambda))}{d\lambda} \\ &= \frac{d}{d\lambda} \left( e^{\frac{n\lambda^2\sigma_H^2(X|Y,\lambda)}{2}} Q(\sqrt{n}\lambda\sigma_H(X|Y,\lambda)) \right) e^{-nr_{X|Y}(\delta(\lambda))} \\ & \quad - e^{\frac{n\lambda^2\sigma_H^2(X|Y,\lambda)}{2}} Q(\sqrt{n}\lambda\sigma_H(X|Y,\lambda)) e^{-nr_{X|Y}(\delta(\lambda))} \frac{d}{d\lambda} (nr_{X|Y}(\delta(\lambda))) \\ &= e^{-nr_{X|Y}(\delta(\lambda))} \left\{ \left[ xe^{\frac{x^2}{2}} Q(x) - \frac{1}{\sqrt{2\pi}} \right] \frac{dx}{d\lambda} - e^{\frac{x^2}{2}} Q(x)n \frac{dr_{X|Y}(\delta)}{d\delta} \Big|_{\delta=\delta(\lambda)} \frac{d\delta(\lambda)}{d\lambda} \right\} \end{aligned} \quad (2.69)$$

where  $x = \sqrt{n}\lambda\sigma_H(X|Y,\lambda)$ . On one hand,

$$\begin{aligned} \frac{dx}{d\lambda} &= \sqrt{n} \left( \sigma_H(X|Y,\lambda) + \lambda \frac{d\sigma_H(X|Y,\lambda)}{d\lambda} \right) \\ &= \sqrt{n} \left( \sigma_H(X|Y,\lambda) + \frac{\lambda}{2\sigma_H(X|Y,\lambda)} \frac{d\sigma_H^2(X|Y,\lambda)}{d\lambda} \right). \end{aligned} \quad (2.70)$$

On the other hand,

$$\frac{dr_{X|Y}(\delta)}{d\delta} \Big|_{\delta=\delta(\lambda)} = \lambda \quad (2.71)$$

$$\frac{d\delta(\lambda)}{d\lambda} = \sigma_H^2(X|Y,\lambda) \quad (2.72)$$

which further implies

$$\begin{aligned} e^{\frac{x^2}{2}} Q(x)n \frac{dr_{X|Y}(\delta)}{d\delta} \Big|_{\delta=\delta(\lambda)} \frac{d\delta(\lambda)}{d\lambda} &= e^{\frac{x^2}{2}} Q(x)n\lambda\sigma_H^2(X|Y,\lambda) \\ &= \sqrt{n}\sigma_H(X|Y,\lambda)xe^{\frac{x^2}{2}} Q(x). \end{aligned} \quad (2.73)$$



Substituting (2.70) and (2.73) into (2.69), we have

$$\begin{aligned}
& \frac{dg_{X|Y,n}(\delta(\lambda))}{d\lambda} \\
&= e^{-nr_{X|Y}(\delta(\lambda))} \left\{ \left[ xe^{\frac{x^2}{2}} Q(x) - \frac{1}{\sqrt{2\pi}} \right] \left( \frac{\sqrt{n}\lambda \frac{d\sigma_H^2(X|Y,\lambda)}{d\lambda}}{2\sigma_H(X|Y,\lambda)} \right) - \frac{\sqrt{n}\sigma_H(X|Y,\lambda)}{\sqrt{2\pi}} \right\} \\
&= e^{-nr_{X|Y}(\delta(\lambda))} \frac{\sqrt{n}\sigma_H(X|Y,\lambda)}{\sqrt{2\pi}} \left\{ \left[ \sqrt{2\pi}xe^{\frac{x^2}{2}} Q(x) - 1 \right] \left( \frac{\lambda \frac{d\sigma_H^2(X|Y,\lambda)}{d\lambda}}{2\sigma_H^2(X|Y,\lambda)} \right) - 1 \right\}. \quad (2.74)
\end{aligned}$$

Note that

$$\begin{aligned}
\sqrt{2\pi}xe^{\frac{x^2}{2}} Q(x) &< \sqrt{2\pi}xe^{\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}x} e^{-\frac{x^2}{2}} \\
&= 1. \quad (2.75)
\end{aligned}$$

If  $\frac{d\sigma_H^2(X|Y,\lambda)}{d\lambda} \geq 0$ , then

$$\left[ \sqrt{2\pi}xe^{\frac{x^2}{2}} Q(x) - 1 \right] \left( \frac{\lambda \frac{d\sigma_H^2(X|Y,\lambda)}{d\lambda}}{2\sigma_H^2(X|Y,\lambda)} \right) \leq 0, \quad (2.76)$$

which further implies that  $\frac{dg_{X|Y,n}(\delta(\lambda))}{d\lambda} < 0$ . In the meantime, if  $\frac{d\sigma_H^2(X|Y,\lambda)}{d\lambda} < 0$ ,

$$\begin{aligned}
& \left[ \sqrt{2\pi}xe^{\frac{x^2}{2}} Q(x) - 1 \right] \left( \frac{\lambda \frac{d\sigma_H^2(X|Y,\lambda)}{d\lambda}}{2\sigma_H^2(X|Y,\lambda)} \right) - 1 \\
&< \left[ \sqrt{2\pi}xe^{\frac{x^2}{2}} \frac{x}{\sqrt{2\pi}(1+x^2)} e^{-\frac{x^2}{2}} - 1 \right] \left( \frac{\lambda \frac{d\sigma_H^2(X|Y,\lambda)}{d\lambda}}{2\sigma_H^2(X|Y,\lambda)} \right) - 1 \\
&= -\frac{1}{1+x^2} \left( \frac{\lambda \frac{d\sigma_H^2(X|Y,\lambda)}{d\lambda}}{2\sigma_H^2(X|Y,\lambda)} \right) - 1 \\
&= -\frac{\lambda \frac{d\sigma_H^2(X|Y,\lambda)}{d\lambda}}{2\sigma_H^2(X|Y,\lambda) (1+n\lambda^2\sigma_H^2(X|Y,\lambda))} - 1. \quad (2.77)
\end{aligned}$$

To continue, let us evaluate  $\frac{d\sigma_H^2(X|Y,\lambda)}{d\lambda}$ . From (2.6), (2.7), and (2.9), it is not hard to verify that

$$\frac{d\sigma_H^2(X|Y,\lambda)}{d\lambda} = \sum_{x \in \mathcal{X}} \int p(x,y) \frac{\partial f_\lambda(x,y)}{\partial \lambda} \ln^2 p(x|y) dy - 2\sigma_H^2(X|Y,\lambda) (H(X) + \delta(\lambda)) \quad (2.78)$$

where

$$\frac{\partial f_\lambda(x,y)}{\partial \lambda} = [-\ln p(x|y) - (H(X|Y) + \delta(\lambda))] f_\lambda(x,y). \quad (2.79)$$

Plugging (2.79) into (2.78) yields

$$\begin{aligned}
& \frac{d\sigma_H^2(X|Y, \lambda)}{d\lambda} \\
&= \mathbf{E} \left( -\ln^3 p(X_\lambda|Y_\lambda) \right) - 3\sigma_H^2(X|Y, \lambda)(H(X|Y) + \delta) - (H(X|Y) + \delta)^3 \\
&= \hat{M}_H(X|Y, \lambda).
\end{aligned} \tag{2.80}$$

Combining (2.74), (2.76), (2.77), and (2.80) together, we have

$$\begin{aligned}
& \frac{dg_{X|Y,n}(\delta(\lambda))}{d\lambda} \\
&\leq e^{-nr_{X|Y}(\delta(\lambda))} \frac{\sqrt{n}\sigma_H(X|Y, \lambda)}{\sqrt{2\pi}} \left( \left| -\frac{\lambda\hat{M}_H(X|Y, \lambda)}{2\sigma_H^2(X|Y, \lambda)(1 + n\lambda^2\sigma_H^2(X|Y, \lambda))} \right| - 1 \right)
\end{aligned} \tag{2.81}$$

$$\leq e^{-nr_{X|Y}(\delta(\lambda))} \frac{\sqrt{n}\sigma_H(X|Y, \lambda)}{\sqrt{2\pi}} \left( \left| -\frac{\lambda\hat{M}_H(X|Y, \lambda)}{2\sigma_H^2(X|Y, \lambda)} \right| - 1 \right) \tag{2.82}$$

In view of the continuity of  $\sigma_H^2(X|Y, \lambda)$  and  $\hat{M}_H(X|Y, \lambda)$  as functions of  $\lambda$ , it is easy to see that there is a  $\lambda^+ > 0$  such that for any  $\lambda \in [0, \lambda^+]$ ,

$$\left| -\frac{\lambda\hat{M}_H(X|Y, \lambda)}{2\sigma_H^2(X|Y, \lambda)} \right| - 1 < 0$$

and hence

$$\frac{dg_{X|Y,n}(\delta(\lambda))}{d\lambda} < 0$$

for any  $n \geq 0$ . This completes the proof of Lemma 1 with  $\delta^+ = \delta(\lambda^+)$ . ■

**Remark 5.** From (2.81), it is clear that when  $n$  is large,

$$\left| -\frac{\lambda\hat{M}_H(X|Y, \lambda)}{2\sigma_H^2(X|Y, \lambda)(1 + n\lambda^2\sigma_H^2(X|Y, \lambda))} \right| - 1 < 0$$

and hence

$$\frac{dg_{X|Y,n}(\delta(\lambda))}{d\lambda} < 0$$

even for  $\lambda \geq \lambda^+$ . Nonetheless, as can be seen later, we are concerned only with the case where  $\delta_n(\epsilon)$  is around 0. Consequently, the exact value of  $\delta^+$  is not important to us.

**Remark 6.** In view of Lemma 1 and the definition of  $\delta_n(\epsilon)$  in (2.67) and (2.68), it follows that  $\delta_n(\frac{1}{2}) = 0$  for any  $n$  and any BIMSC. However, when  $\epsilon < 1/2$ ,  $\delta_n(\epsilon)$  depends not only on  $n$  and  $\epsilon$ , but also on the BIMSC itself through the function  $r_{X|Y}(\delta)$ . Given  $n$  and  $\epsilon < 1/2$ , the value of

$\delta_n(\epsilon)$  fluctuates a lot from one BIMSC to another through the behavior of  $r_{X|Y}(\delta)$  around  $\delta = 0$ , which depends on both the second and third order derivatives of  $r_{X|Y}(\delta)$ . Given a BIMSC, if  $r_{X|Y}(\delta)$  is approximated as in (2.16), then  $\delta_n(\epsilon)$  is in the order of  $\sqrt{\frac{-\ln \epsilon}{n}}$ . Of course, such an approximation is accurate only when  $\delta$  or  $\sqrt{\frac{-\ln \epsilon}{n}}$  is sufficiently small.

With respect to  $\delta_n(\epsilon)$ ,  $R_n(\epsilon)$  has a nice Taylor-type expansion, as shown in Theorem 2.

**Theorem 2.** Given a BIMSC, for any  $n$  and  $\epsilon$  satisfying  $g_{X|Y,n}(\delta^+/2) \leq \epsilon < 1$ ,

$$|R_n(\epsilon) - (C_{\text{BIMSC}} - \delta_n(\epsilon))| \leq o(\delta_n(\epsilon)) \quad (2.83)$$

where

$$o(\delta_n(\epsilon)) = r_{X|Y}(\delta_n(\epsilon)) + \frac{\ln n + d_1}{n} \quad (2.84)$$

if  $\epsilon \leq \frac{1}{3}$ , and

$$\left| R_n(\epsilon) - \left( C_{\text{BIMSC}} - \frac{\sigma_H(X|Y)}{\sqrt{n}} Q^{-1}(\epsilon) \right) \right| \leq \frac{\ln n + d_2}{n} \quad (2.85)$$

otherwise, where  $d_1$  and  $d_2$  are channel parameters independent of both  $n$  and  $\epsilon$ .

*Proof:* When  $\epsilon > \frac{1}{3}$ , (2.85) can be easily proved by combining (1.5), (1.6) and (2.34). Therefore, it suffices for us to show (2.83) and (2.84) for  $\epsilon \leq \frac{1}{3}$ . By (1.1) and definition of  $\bar{\xi}_H(X|Y, \lambda, n)$ , for any BIMSC there exists a channel code  $\mathcal{C}_n$  such that

$$\begin{aligned} P_e(\mathcal{C}_n) &\leq \left( \bar{\xi}_H(X|Y, \lambda, n) + \frac{2(1 - C_{BE})M_H(X|Y, \lambda)}{\sqrt{n}\sigma_H^3(X|Y, \lambda)} \right) e^{-nr_{X|Y}(\delta)} \\ &\leq g_{X|Y,n}(\delta) + \frac{2M_H(X|Y, \lambda)}{\sqrt{n}\sigma_H^3(X|Y, \lambda)} e^{-nr_{X|Y}(\delta)} \end{aligned} \quad (2.86)$$

and

$$R(\mathcal{C}_n) \geq C_{\text{BIMSC}} - \delta + \frac{\ln \left[ \frac{2(1 - C_{BE})M_H(X|Y, \lambda)}{\sqrt{n}\sigma_H^3(X|Y, \lambda)} e^{-nr_{X|Y}(\delta)} \right]}{n} \quad (2.87)$$

which implies that for any  $\delta$  such that

$$g_{X|Y,n}(\delta) + \frac{2M_H(X|Y, \lambda)}{\sqrt{n}\sigma_H^3(X|Y, \lambda)} e^{-nr_{X|Y}(\delta)} \leq \epsilon \quad (2.88)$$

the following inequality holds

$$R_n(\epsilon) \geq C_{\text{BIMSC}} - \delta + \frac{\ln \left[ \frac{2(1 - C_{BE})M_H(X|Y, \lambda)}{\sqrt{n}\sigma_H^3(X|Y, \lambda)} e^{-nr_{X|Y}(\delta)} \right]}{n} \quad (2.89)$$

where  $\lambda = r'_{X|Y}(\delta)$ . Now let  $\bar{\delta} = \delta_n(\epsilon) + \frac{\eta}{n}$  for some constant  $\eta > 0$ , which will be specified later, and  $\bar{\lambda} = r'_{X|Y}(\bar{\delta})$ . By convexity of  $r_{X|Y}(\delta)$ ,

$$r_{X|Y}(\bar{\delta}) \geq r_{X|Y}(\delta_n(\epsilon)) + \lambda_n(\epsilon) \frac{\eta}{n} \quad (2.90)$$

where  $\lambda_n(\epsilon) = r'_{X|Y}(\delta_n(\epsilon))$ . Then

$$\begin{aligned} & g_{X|Y,n}(\bar{\delta}) + \frac{2M_H(X|Y, \bar{\lambda})}{\sqrt{n}\sigma_H^3(X|Y, \bar{\lambda})} e^{-nr_{X|Y}(\bar{\delta})} \\ & \stackrel{1)}{\leq} \left( e^{\frac{n\bar{\lambda}^2\sigma_H^2(X|Y, \bar{\lambda})}{2}} Q(\sqrt{n}\bar{\lambda}\sigma_H(X|Y, \bar{\lambda})) + \frac{2M_H(X|Y, \bar{\lambda})}{\sqrt{n}\sigma_H^3(X|Y, \bar{\lambda})} \right) e^{-n(r_{X|Y}(\delta_n(\epsilon)) + \lambda_n(\epsilon)\frac{\eta}{n})} \\ & = \left( 1 + \frac{\frac{2M_H(X|Y, \bar{\lambda})}{\sqrt{n}\sigma_H^3(X|Y, \bar{\lambda})}}{e^{\frac{n\bar{\lambda}^2\sigma_H^2(X|Y, \bar{\lambda})}{2}} Q(\sqrt{n}\bar{\lambda}\sigma_H(X|Y, \bar{\lambda}))} \right) e^{\frac{n\bar{\lambda}^2\sigma_H^2(X|Y, \bar{\lambda})}{2}} Q(\sqrt{n}\bar{\lambda}\sigma_H(X|Y, \bar{\lambda})) \\ & \quad \times e^{-nr_{X|Y}(\delta_n(\epsilon)) - \eta\lambda_n(\epsilon)} \\ & \stackrel{2)}{\leq} \left( 1 + \frac{2M_H(X|Y, \bar{\lambda})\sqrt{2\pi}\bar{\lambda} \left( 1 + \frac{1}{n\bar{\lambda}^2\sigma_H^2(X|Y, \bar{\lambda})} \right)}{\sigma_H^2(X|Y, \bar{\lambda})} \right) \\ & \quad \times e^{\frac{n\lambda_n^2(\epsilon)\sigma_H^2(X|Y, \lambda_n(\epsilon))}{2}} Q(\sqrt{n}\lambda_n(\epsilon)\sigma_H(X|Y, \lambda_n(\epsilon))) e^{-nr_{X|Y}(\delta_n(\epsilon)) - \eta\lambda_n(\epsilon)} \\ & = g_{X|Y,n}(\delta_n(\epsilon)) e^{-\eta\lambda_n(\epsilon)} \left( 1 + \frac{2\sqrt{2\pi}M_H(X|Y, \bar{\lambda}) \left( 1 + \frac{1}{n\bar{\lambda}^2\sigma_H^2(X|Y, \bar{\lambda})} \right)}{\sigma_H^2(X|Y, \bar{\lambda})} \bar{\lambda} \right) \\ & \stackrel{3)}{=} \epsilon e^{-\eta\lambda_n(\epsilon)} \left( 1 + \frac{2\sqrt{2\pi}M_H(X|Y, \bar{\lambda}) \left( 1 + \frac{1}{n\bar{\lambda}^2\sigma_H^2(X|Y, \bar{\lambda})} \right)}{\sigma_H^2(X|Y, \bar{\lambda})} \left( \lambda_n(\epsilon) + \frac{1}{\sigma_H^2(X|Y, \bar{\lambda})} \frac{\eta}{n} \right) \right) \\ & \stackrel{4)}{\leq} \epsilon \frac{1 + \frac{2\sqrt{2\pi}M_H(X|Y, \bar{\lambda}) \left( 1 + \frac{1}{n\bar{\lambda}^2\sigma_H^2(X|Y, \bar{\lambda})} \right)}{\sigma_H^2(X|Y, \bar{\lambda})} \left( \lambda_n(\epsilon) + \frac{1}{\sigma_H^2(X|Y, \bar{\lambda})} \frac{\eta}{n} \right)}{1 + \eta\lambda_n(\epsilon) + \frac{1}{2}\eta^2\lambda_n^2(\epsilon)}. \end{aligned} \quad (2.91)$$

In the derivation of (2.91), the inequality 1) is due to (2.90); the inequality 2) follows from the fact that  $e^{\frac{x^2}{2}} Q(x)$  is a strictly decreasing function of  $x$ ,  $\lambda\sigma_H(X|Y, \lambda)$  is strictly increasing with respect to  $\lambda$  as shown below

$$\begin{aligned} \frac{d\lambda\sigma_H(X|Y, \lambda)}{d\lambda} &= \sigma_H(X|Y, \lambda) + \lambda \frac{d\sigma_H(X|Y, \lambda)}{d\lambda} \\ &= \sigma_H(X|Y, \lambda) \left( 1 + \lambda \frac{\frac{d\sigma_H^2(X|Y, \lambda)}{d\lambda}}{2\sigma_H^2(X|Y, \lambda)} \right) \\ &= \sigma_H(X|Y, \lambda) \left( 1 + \lambda \frac{\hat{M}_H(X|Y, \lambda)}{2\sigma_H^2(X|Y, \lambda)} \right) \\ &> 0 \end{aligned} \quad (2.92)$$

for  $\lambda \in [0, \lambda^+]$ , and

$$e^{\frac{x^2}{2}} Q(x) \geq \frac{x}{\sqrt{2\pi}(1+x^2)}; \quad (2.93)$$

the equality 3) is attributable to

$$\bar{\lambda} = \lambda_n(\epsilon) + \frac{d\lambda}{d\delta} \bigg|_{\lambda=\bar{\lambda}} \frac{\eta}{n} = \lambda_n(\epsilon) + \frac{1}{\sigma_H^2(X|Y, \bar{\lambda})} \frac{\eta}{n} \quad (2.94)$$

for some  $\tilde{\lambda} \in [\lambda_n(\epsilon), \bar{\lambda}]$ ; and finally, the inequality 4) follows from the inequality

$$e^x > 1 + x + \frac{x^2}{2}$$

for any  $x > 0$ . In order to satisfy (2.88), let us now choose  $\eta$  such that

$$\eta \lambda_n(\epsilon) \geq \frac{2\sqrt{2\pi} M_H(X|Y, \bar{\lambda}) \left(1 + \frac{1}{n\bar{\lambda}^2 \sigma_H^2(X|Y, \bar{\lambda})}\right)}{\sigma_H^2(X|Y, \bar{\lambda})} \lambda_n(\epsilon) \quad (2.95)$$

and

$$\frac{1}{2} \eta^2 \lambda_n^2(\epsilon) \geq \frac{2\sqrt{2\pi} M_H(X|Y, \bar{\lambda}) \left(1 + \frac{1}{n\bar{\lambda}^2 \sigma_H^2(X|Y, \bar{\lambda})}\right)}{\sigma_H^2(X|Y, \bar{\lambda})} \frac{1}{\sigma_H^2(X|Y, \tilde{\lambda})} \frac{\eta}{n}, \quad (2.96)$$

i.e.

$$\eta = \frac{2\sqrt{2\pi} M_H(X|Y, \bar{\lambda}) \left(1 + \frac{1}{n\bar{\lambda}^2 \sigma_H^2(X|Y, \bar{\lambda})}\right)}{\sigma_H^2(X|Y, \bar{\lambda})} \max \left\{ 1, \frac{2}{n\lambda_n^2(\epsilon) \sigma_H^2(X|Y, \tilde{\lambda})} \right\}. \quad (2.97)$$

To see  $\eta$  is bounded, note that  $\frac{M_H(X|Y, \lambda)}{\sigma_H^2(X|Y, \lambda)}$  is always bounded for  $\lambda \in [0, \lambda^+]$ . On the other hand, for  $\epsilon \leq \frac{1}{3}$ ,  $\sqrt{n} \lambda_n(\epsilon) \sigma_H(X|Y, \lambda_n(\epsilon)) > c$  for some constant  $c$ , as  $\sqrt{n} \lambda_n(\epsilon) \sigma_H(X|Y, \lambda_n(\epsilon)) \rightarrow 0$  implies that  $\epsilon = g_{X|Y, n}(\delta_n(\epsilon)) \rightarrow \frac{1}{2}$ , and the same argument can be applied to  $\sqrt{n} \lambda_n(\epsilon) \sigma_H^2(X|Y, \tilde{\lambda})$ .

Therefore,

$$\eta \leq 2\sqrt{2\pi} \max_{\lambda \in [0, \lambda^+]} \left[ \frac{M_H(X|Y, \lambda)}{\sigma_H^2(X|Y, \lambda)} \right] (1 + c^{-2}) \max \{1, 2c^{-2}\}. \quad (2.98)$$

Then combining (2.88), (2.89), (2.90), (2.91), (2.95) and (2.96) yields

$$\begin{aligned}
R_n(\epsilon) &\geq C_{\text{BIMSC}} - \bar{\delta} + \frac{\ln \left[ \frac{2(1-C_{BE})M_H(X|Y, \bar{\lambda})}{\sqrt{n}\sigma_H^3(X|Y, \bar{\lambda})} e^{-nr_{X|Y}(\bar{\delta})} \right]}{n} \\
&= C_{\text{BIMSC}} - \bar{\delta} - r_{X|Y}(\bar{\delta}) + \frac{\ln \left[ \frac{2(1-C_{BE})M_H(X|Y, \bar{\lambda})}{\sigma_H^3(X|Y, \bar{\lambda})} \right] - \frac{1}{2} \ln n}{n} \\
&\stackrel{1)}{\geq} C_{\text{BIMSC}} - \delta_n(\epsilon) - r_{X|Y}(\delta_n(\epsilon)) - \bar{\lambda} \frac{\eta}{n} + \frac{\ln \left[ \frac{2(1-C_{BE})M_H(X|Y, \bar{\lambda})}{\sigma_H^3(X|Y, \bar{\lambda})} \right] - \eta - \frac{1}{2} \ln n}{n} \\
&\geq C_{\text{BIMSC}} - \delta_n(\epsilon) - r_{X|Y}(\delta_n(\epsilon)) \\
&\quad + \frac{-\lambda^+ \eta + \ln \left[ 2(1-C_{BE}) \min_{\lambda} \left( \frac{2M_H(X|Y, \lambda)}{\sigma_H^3(X|Y, \lambda)} \right) \right] - \eta - \frac{1}{2} \ln n}{n} \\
&= C_{\text{BIMSC}} - \delta_n(\epsilon) - r_{X|Y}(\delta_n(\epsilon)) - \frac{\frac{1}{2} \ln n + \bar{d}_1}{n}, \tag{2.99}
\end{aligned}$$

where  $\bar{d}_1$  is independent of both  $n$  and  $\epsilon$ . In the derivation of (2.99), the inequality 1) follows from the convexity of  $r_{X|Y}(\delta)$  and the fact that

$$r_{X|Y}(\bar{\delta}) \leq r_{X|Y}(\delta_n(\epsilon)) + \bar{\lambda} \frac{\eta}{n}.$$

We now proceed to establish an upper bound on  $R_n(\epsilon)$ . Towards this end, recall (2.30) and (2.31) where we make a small modification by choosing  $\beta_n = \lambda = r'_{X|Y}(\delta)$  in the proof of Theorem 1. Then for any  $\delta$  such that

$$(1 + 2\lambda)\epsilon \leq \underline{\xi}_H(X|Y, \lambda, n) e^{-nr_{X|Y}(\delta)} \tag{2.100}$$

we have

$$\begin{aligned}
R_n(\epsilon) &\leq C_{\text{BIMSC}} - \delta - \frac{\ln \epsilon - \ln P(B_{n,\delta}) + 2 \ln \lambda - \ln(1 + \lambda)}{n} \\
&\leq C_{\text{BIMSC}} - \delta + \frac{-\ln \epsilon - 2 \ln \lambda + \lambda}{n} \tag{2.101}
\end{aligned}$$

where the trivial bound  $P(B_{n,\delta}) \leq 1$  is applied. Now let  $\underline{\delta} = \delta_n(\epsilon) - \frac{\eta'}{n}$  for some constant  $\eta' > 0$ ,

which will be specified later, and  $\underline{\lambda} = r'_{X|Y}(\underline{\delta})$ . Then

$$\begin{aligned}
& \xi_H(X|Y, \underline{\lambda}, n) e^{-nr_{X|Y}(\underline{\delta})} \\
& \stackrel{1)}{\geq} e^{\frac{n\lambda^2\sigma_H^2(X|Y, \underline{\lambda})}{2}} Q(\rho_* + \sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda})) e^{-nr_{X|Y}(\delta_n(\epsilon)) + \underline{\lambda}\eta'} \\
& = e^{\frac{n\lambda^2\sigma_H^2(X|Y, \underline{\lambda})}{2}} Q(\sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda})) \frac{Q(\rho_* + \sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda}))}{Q(\sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda}))} e^{-nr_{X|Y}(\delta_n(\epsilon)) + \underline{\lambda}\eta'} \\
& \stackrel{2)}{\geq} g_{X|Y, n}(\delta_n(\epsilon)) \frac{Q(\rho_* + \sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda}))}{Q(\sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda}))} e^{\underline{\lambda}\eta'} \\
& \stackrel{3)}{\geq} (1 + 2\underline{\lambda})\epsilon \frac{Q(\rho_* + \sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda}))}{Q(\sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda}))} e^{\underline{\lambda}(\eta' - 2)}. \tag{2.102}
\end{aligned}$$

In the derivation of (2.102), the inequality 1) is due to the convexity of  $r_{X|Y}(\delta)$  and the fact that

$$r_{X|Y}(\underline{\delta}) \leq r_{X|Y}(\delta_n(\epsilon)) - \underline{\lambda} \frac{\eta'}{n};$$

the inequality 2) follows again from the fact that  $e^{\frac{x^2}{2}} Q(x)$  is a strictly decreasing function of  $x$  and  $\lambda\sigma_H(X|Y, \lambda)$  is increasing with respect to  $\lambda$ ; and finally the inequality 3) is attributable to the inequality  $e^x \geq 1 + x$  for any  $x \geq 0$ .

In order for (2.100) to be satisfied, we now choose  $\eta'$  such that

$$\begin{aligned}
\eta' &= 2 + \frac{1}{\underline{\lambda}} \ln \frac{Q(\sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda}))}{Q(\rho_* + \sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda}))} \\
&= 2 + \frac{1}{\underline{\lambda}} \ln \left[ 1 + \rho_* \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{(\tilde{\rho} + \sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda}))^2}{2}}}{Q(\rho_* + \sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda}))} \right] \tag{2.103}
\end{aligned}$$

where  $0 \leq \tilde{\rho} \leq \rho_*$ . One can verify that

$$\begin{aligned}
\eta' &\leq 2 + \frac{\rho_*}{\underline{\lambda}} \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{(\tilde{\rho} + \sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda}))^2}{2}}}{Q(\rho_* + \sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda}))} \\
&\leq 2 + \frac{\rho_*}{\underline{\lambda}} \frac{1 + (\rho_* + \sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda}))^2}{\rho_* + \sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda})} e^{\sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda})(\rho_* - \tilde{\rho}) + \frac{\rho_*^2 - \tilde{\rho}^2}{2}} \tag{2.104}
\end{aligned}$$

where the last inequality is due to (2.93). From the definition of  $\rho_*$ , it is not hard to see that  $\rho_* = \frac{\eta''}{\sqrt{n}}$  for some constant  $\eta''$  depending only on channel parameters. Meanwhile, we have  $\sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda}) > c$  as discussed above. Then

$$\begin{aligned}
\eta' &\leq 2 + \frac{\eta''}{\sqrt{n}\underline{\lambda}} \left( c^{-1} + \frac{\eta''}{\sqrt{n}} + \sqrt{n}\underline{\lambda}\sigma_H(X|Y, \underline{\lambda}) \right) e^{\eta''\lambda^+ \max_{\lambda \in [0, \lambda^+]} \sigma_H(X|Y, \lambda) + \frac{(\eta'')^2}{2n}} \\
&\leq 2 + (c^{-2} + c^{-1}\eta'' + 1) \eta'' \left[ \max_{\lambda \in [0, \lambda^+]} \sigma_H(X|Y, \lambda) \right] e^{\eta''\lambda^+ \max_{\lambda \in [0, \lambda^+]} \sigma_H(X|Y, \lambda) + (\eta'')^2} \tag{2.105}
\end{aligned}$$

which is independent of both  $n$  and  $\epsilon$ . Now combining (2.102) and (2.103), we have

$$\xi_H(X|Y, \underline{\lambda}, n) e^{-nr_{X|Y}(\underline{\delta})} \geq (1 + 2\underline{\lambda})\epsilon \quad (2.106)$$

and consequently,

$$\begin{aligned} R_n(\epsilon) &\leq C_{\text{BIMSC}} - \underline{\delta} + \frac{-\ln \epsilon - 2 \ln \underline{\lambda} + \underline{\lambda}}{n} \\ &\stackrel{1)}{\leq} C_{\text{BIMSC}} - \delta_n(\epsilon) + r_{X|Y}(\delta_n(\epsilon)) \\ &\quad + \frac{\ln \left[ \sqrt{2\pi} \sqrt{n} \lambda_n(\epsilon) \sigma_H(X|Y, \lambda_n(\epsilon)) \left( 1 + \frac{1}{n \lambda_n^2(\epsilon) \sigma_H^2(X|Y, \lambda_n(\epsilon))} \right) \right]}{n} \\ &\quad + \frac{-2 \ln \underline{\lambda} + \lambda^+ + \eta'}{n} \\ &= C_{\text{BIMSC}} - \delta_n(\epsilon) + r_{X|Y}(\delta_n(\epsilon)) + \frac{\ln \left( 1 + \frac{1}{n \lambda_n^2(\epsilon) \sigma_H^2(X|Y, \lambda_n(\epsilon))} \right)}{n} \\ &\quad + \frac{\ln n + \ln \sqrt{2\pi} \sigma_H(X|Y, \lambda_n(\epsilon)) + \ln \frac{\lambda_n(\epsilon)}{\underline{\lambda}} - \ln \sqrt{n} \underline{\lambda} + \lambda^+ + \eta'}{n} \\ &\stackrel{2)}{\leq} C_{\text{BIMSC}} - \delta_n(\epsilon) + r_{X|Y}(\delta_n(\epsilon)) + \frac{\ln n + \underline{d}_1}{n} \end{aligned} \quad (2.107)$$

where  $\underline{d}_1$  is another constant depending only on the channel. In the derivation of (2.107), the inequality 1) is due to (2.93) and the definition of  $\delta_n(\epsilon)$  in (2.68); and the inequality 2) follows from the fact that

$$\frac{\lambda_n(\epsilon)}{\underline{\lambda}} = 1 + \frac{1}{\sigma_H^2(X|Y, \hat{\lambda})} \frac{\eta'}{n \underline{\lambda}}$$

for some  $\hat{\lambda} \in [\underline{\lambda}, \lambda_n(\epsilon)]$  and

$$\sqrt{n} \underline{\lambda} \sigma_H(X|Y, \underline{\lambda}) > c.$$

Then the theorem is proved by combining (2.99) and (2.107) and making  $d_1 = \max\{\bar{d}_1, \underline{d}_1\}$ . ■

**Remark 7.** The condition  $\epsilon \leq \frac{1}{3}$  for (2.83) and (2.84) can be relaxed as we only require that  $\sqrt{n} \delta_n(\epsilon)$  or equivalently  $\sqrt{n} \lambda$  be lower bounded by a constant, which is true when  $\epsilon \leq d$  for any constant  $d < \frac{1}{2}$ . In addition, when  $\epsilon \leq g_{X|Y,n}(\delta^+/2)$ ,  $\epsilon$  is an exponential function of  $n$ , in which case the maximum achievable rate is below the channel capacity by a positive constant even when  $n$  goes to  $\infty$ . As such, from a practical point of view, the case  $\epsilon \leq g_{X|Y,n}(\delta^+/2)$  is not interesting, especially when one can approach the channel capacity very closely as shown in the achievability given in (1.1) and (1.2).



**Remark 8.** In the definition of  $R_n(\epsilon)$ , the average error probability is used. If the maximal error probability is used instead, Theorem 2 remains valid. This can be proved similarly by first using the standard technique of removing bad codewords from the code in the achievability given in (1.1) and (1.2) to establish similar achievability with maximal error probability and then combining it with Corollary 1.

**Remark 9.** In view of Theorem 2, it is now clear that jar decoding is indeed optimal up to the second order coding performance in the non-asymptotical regime. Since the achievability given in (1.1) and (1.2) was established for linear block codes, it follows from Theorem 2 that linear block coding is also optimal up to the second order coding performance in the non-asymptotical regime for any BIMSC. In addition, in the Taylor-type expansion of  $R_n(\epsilon)$ , the third order term is  $O(\delta_n^2(\epsilon))$  whenever  $\delta_n(\epsilon) = \Omega(\sqrt{\ln n/n})$  since it follows from (2.16) that  $r_{X|Y}(\delta_n(\epsilon)) = O(\delta_n^2(\epsilon))$ .

#### D. Comparison with Asymptotic Analysis

It is instructive to compare Theorem 2 with the second order asymptotic performance analysis as  $n$  goes to  $\infty$ .

*Asymptotic analysis with constant  $0 < \epsilon < 1$  and  $n \rightarrow \infty$ :* Fix  $0 < \epsilon < 1$ . It was shown in [3], [4], [5] that for a BIMSC with a discrete output alphabet

$$R_n(\epsilon) = C_{\text{BIMSC}} - \frac{\sigma_H(X|Y)}{\sqrt{n}} Q^{-1}(\epsilon) + O\left(\frac{\ln n}{n}\right) \quad (2.108)$$

for sufficiently large  $n$ . The expression  $C_{\text{BIMSC}} - \frac{\sigma_H(X|Y)}{\sqrt{n}} Q^{-1}(\epsilon)$  was referred to as the normal approximation for  $R_n(\epsilon)$ . Clearly, when  $\epsilon > 1/3$ , (2.108) is essentially the same as (2.85). Let us now look at the case  $\epsilon \leq 1/3$ . In this case, by using the Taylor expansion of  $r_{X|Y}(\delta)$  around  $\delta = 0$

$$\begin{aligned} r_{X|Y}(\delta) &= \frac{1}{2\sigma_H^2(X|Y)} \delta^2 + \frac{\left. \frac{d\sigma_H^2(X|Y, \lambda)}{d\lambda} \right|_{\lambda=0}}{6\sigma_H^6(X|Y)} \delta^3 + O(\delta^4) \\ &= \frac{1}{2\sigma_H^2(X|Y)} \delta^2 + \frac{-\hat{M}_H(X|Y)}{6\sigma_H^6(X|Y)} \delta^3 + O(\delta^4) \end{aligned} \quad (2.109)$$

it can be verified that

$$\delta_n(\epsilon) = \frac{\sigma_H(X|Y)}{\sqrt{n}} Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right). \quad (2.110)$$

Thus the Taylor-type expansion of  $R_n(\epsilon)$  in Theorem 2 implies the second order asymptotic analysis with constant  $0 < \epsilon < 1$  and  $n \rightarrow \infty$  shown in (2.108).

*Asymptotic analysis with  $n \rightarrow \infty$  and non-exponentially decaying  $\epsilon$ :* Suppose now  $\epsilon$  is a function of  $n$  and goes to 0 as  $n \rightarrow \infty$ , but at a non-exponential speed. In this case, as  $n \rightarrow \infty$ ,  $\delta_n(\epsilon)$  goes to 0 at the speed of  $\Theta\left(\sqrt{\frac{-\ln \epsilon}{n}}\right)$ , and  $\sqrt{n}\lambda_n(\epsilon)$  goes to  $\infty$ . By ignoring the third and higher order terms in the Taylor expansion of  $r_{X|Y}(\delta)$ , one has the following approximations:

$$g_{X|Y,n}(\delta_n(\epsilon)) \approx \frac{1}{\sqrt{2\pi}\sqrt{n}\lambda_n(\epsilon)\sigma_H(X|Y, \lambda_n(\epsilon))} e^{-n\frac{\delta_n^2(\epsilon)}{2\sigma_H^2(X|Y)}} \quad (2.111)$$

and

$$Q(x) \approx \frac{1}{\sqrt{2\pi}x} e^{-\frac{x^2}{2}} \text{ for large } x.$$

By these approximations, it is not hard to verify that in this case

$$\lim_{n \rightarrow \infty} \frac{\delta_n(\epsilon)}{\frac{\sigma_H(X|Y)}{\sqrt{n}}Q^{-1}(\epsilon)} = 1.$$

Therefore, from Theorem 2, it follows that when  $\epsilon$  goes to 0 at a non-exponential speed as  $n \rightarrow \infty$ ,  $\frac{\sigma_H(X|Y)}{\sqrt{n}}Q^{-1}(\epsilon)$  is still the second order term of  $R_n(\epsilon)$  in the asymptotic analysis with  $n \rightarrow \infty$ . Indeed, this can also be verified by looking at the specific case given by (1.3), (1.4), and (2.33) when  $\epsilon$  goes to 0 at a polynomial speed as  $n \rightarrow \infty$ . To the best of our knowledge, the second order asymptotic analysis with  $n \rightarrow \infty$  and non-exponentially decaying  $\epsilon$  has not been addressed before in the literature.

*Divergence of  $\delta_n(\epsilon)$  from  $\frac{\sigma_H(X|Y)}{\sqrt{n}}Q^{-1}(\epsilon)$ :* The agreement between  $\delta_n(\epsilon)$  and  $\frac{\sigma_H(X|Y)}{\sqrt{n}}Q^{-1}(\epsilon)$  terminates when the third order term

$$\frac{-\hat{M}_H(X|Y)}{6\sigma_H^6(X|Y)}\delta^3$$

in the Taylor expansion of  $r_{X|Y}(\delta)$  shown in (2.109) can not be ignored. This happens when  $\delta$  is not small, which is typical in practice for finite block length  $n$ , or

$$\zeta_{X|Y} \triangleq \frac{-\hat{M}_H(X|Y)}{6\sigma_H^6(X|Y)} \quad (2.112)$$

is large. In this case,  $\frac{\sigma_H(X|Y)}{\sqrt{n}}Q^{-1}(\epsilon)$  will be smaller than  $\delta_n(\epsilon)$  by a relatively large margin if  $\zeta_{X|Y} < 0$ , and larger than  $\delta_n(\epsilon)$  by a relatively large margin if  $\zeta_{X|Y} > 0$ . As such, the normal approximation would fail to provide a reasonable estimate for  $R_n(\epsilon)$ . This will be further confirmed by numerical results shown in Section IV for well known channels such as the BEC, BSC, and BIAGN for finite  $n$ .

### III. NON-ASYMPTOTIC CONVERSE AND TAYLOR-TYPE EXPANSION: DIMC

We now extend Theorems 1 and 2 to the case of DIMC  $P = \{p(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$ , where  $\mathcal{X}$  is discrete, but  $\mathcal{Y}$  is arbitrary (discrete or continuous).

#### A. Definitions

Let  $\mathcal{P}$  denote the set of all distributions over  $\mathcal{X}$ . Let  $\mathcal{P}_n$  denote the set of types on  $\mathcal{X}^n$  with denominator  $n$  [7], and  $t(x^n)$  be the type of  $x^n$ . Moreover, for  $t \in \mathcal{P}_n$ , let

$$\mathcal{X}_t^n \triangleq \{x^n \in \mathcal{X}^n : t(x^n) = t\}. \quad (3.1)$$

Before stating our converse channel coding theorem for DIMC, we again need to introduce some definitions from [6]. For any  $t \in \mathcal{P}$ , define

$$q_t(y^n) \triangleq \prod_{i=1}^n q_t(y_i) \quad (3.2)$$

where

$$q_t(y) \triangleq \sum_{x \in \mathcal{X}} t(x) p(y|x), \quad (3.3)$$

$$I(t; P) \triangleq \sum_{x \in \mathcal{X}} t(x) \int p(y|x) \ln \frac{p(y|x)}{q_t(y)} dy \quad (3.4)$$

and

$$\lambda_-^*(t; P) \triangleq \sup \left\{ \lambda \geq 0 : \sum_{a \in \mathcal{X}} t(a) \int p(y|a) \left[ \frac{p(y|a)}{q_t(y)} \right]^{-\lambda} dy < \infty \right\}. \quad (3.5)$$

It is easy to see that  $\lambda_-^*(t; P)$  is the same for all  $t \in \mathcal{P}$  with the same support set  $\{a \in \mathcal{X} : t(a) > 0\}$ . Suppose that

$$\lambda_-^*(t; P) > 0. \quad (3.6)$$

Define for any  $t \in \mathcal{P}$  and any  $\delta \geq 0$

$$r_-(t, \delta) \triangleq \sup_{\lambda \geq 0} \left[ \lambda(\delta - I(t; P)) - \sum_{x \in \mathcal{X}} t(x) \ln \int p(y|x) \left[ \frac{p(y|x)}{q_t(y)} \right]^{-\lambda} dy \right] \quad (3.7)$$

and for any  $t \in \mathcal{P}$  and any  $\lambda \in [0, \lambda_-^*(t; P))$ , random variables  $X_t$  and  $Y_{t,\lambda}$  with joint distribution  $t(x)p(y|x)f_{-\lambda}(y|x)$  where

$$f_{-\lambda}(y|x) \triangleq \frac{\left[ \frac{p(y|x)}{q_t(y)} \right]^{-\lambda}}{\int p(v|x) \left[ \frac{p(v|x)}{q_t(v)} \right]^{-\lambda} dv}. \quad (3.8)$$

Then define

$$D(t, x, \lambda) \triangleq \mathbf{E} \left[ \ln \frac{p(Y_{t,\lambda}|X_t)}{q_t(Y_{t,\lambda})} \middle| X_t = x \right] \quad (3.9)$$

$$\delta_-(t, \lambda) \triangleq \mathbf{E} \left[ -\ln \frac{p(Y_{t,\lambda}|X_t)}{q_t(Y_{t,\lambda})} \right] + I(t; P) \quad (3.10)$$

$$\Delta_-^*(t) \triangleq \lim_{\lambda \uparrow \lambda_-^*(t; P)} \delta_-(t, \lambda) \quad (3.11)$$

$$\begin{aligned} \sigma_{D,-}^2(t; P, \lambda) &\triangleq \mathbf{E} \left\{ \mathbf{Var} \left[ \ln \frac{p(Y_{t,\lambda}|X_t)}{q_t(Y_{t,\lambda})} \middle| X_t \right] \right\} \\ &= \sum_{x \in \mathcal{X}} t(x) \mathbf{Var} \left[ \ln \frac{p(Y_{t,\lambda}|X_t)}{q_t(Y_{t,\lambda})} \middle| X_t = x \right] \end{aligned} \quad (3.12)$$

$$\begin{aligned} M_{D,-}(t; P, \lambda) &\triangleq \mathbf{E} \left\{ \mathbf{M}_3 \left[ \ln \frac{p(Y_{t,\lambda}|X_t)}{q_t(Y_{t,\lambda})} \middle| X_t \right] \right\} \\ &= \sum_{x \in \mathcal{X}} t(x) \mathbf{M}_3 \left[ \ln \frac{p(Y_{t,\lambda}|X_t)}{q_t(Y_{t,\lambda})} \middle| X_t = x \right] \end{aligned} \quad (3.13)$$

and

$$\begin{aligned} \hat{M}_{D,-}(t; P, \lambda) &\triangleq \mathbf{E} \left\{ \hat{\mathbf{M}}_3 \left[ \ln \frac{p(Y_{t,\lambda}|X_t)}{q_t(Y_{t,\lambda})} \middle| X_t \right] \right\} \\ &= \sum_{x \in \mathcal{X}} t(x) \hat{\mathbf{M}}_3 \left[ \ln \frac{p(Y_{t,\lambda}|X_t)}{q_t(Y_{t,\lambda})} \middle| X_t = x \right]. \end{aligned} \quad (3.14)$$

Note that  $\sigma_{D,-}^2(t; P, \lambda)$ ,  $M_{D,-}(t; P, \lambda)$ , and  $\hat{M}_{D,-}(t; P, \lambda)$  are respectively the conditional variance, conditional third absolute central moment, and conditional third central moment of  $\ln \frac{p(Y_{t,\lambda}|X_t)}{q_t(Y_{t,\lambda})}$  given  $X_t$ . Write  $\sigma_{D,-}^2(t; P, 0)$  simply as  $\sigma_D^2(t; P)$ ,  $M_{D,-}(t; P, 0)$  as  $M_D(t; P)$ , and  $\hat{M}_{D,-}(t; P, 0)$  as  $\hat{M}_D(t; P)$ . Assume that

$$\sigma_D^2(t; P) > 0 \text{ and } M_D(t; P) < \infty. \quad (3.15)$$

Furthermore  $r_-(t, \delta)$  has the following parametric expression

$$r_-(t, \delta_-(t, \lambda)) = \lambda(\delta_-(t, \lambda) - I(t; P)) - \sum_{x \in \mathcal{X}} t(x) \ln \int p(y|x) \left[ \frac{p(y|x)}{q_t(y)} \right]^{-\lambda} dy \quad (3.16)$$

with  $\lambda = \frac{\partial r_-(t, \delta)}{\partial \delta}$  satisfying  $\delta_-(t, \lambda) = \delta$ . In addition, let

$$\begin{aligned} \bar{\xi}_{D,-}(t; P, \lambda, n) &\triangleq \frac{2C_{BE}M_{D,-}(t; P, \lambda)}{\sqrt{n}\sigma_{D,-}^3(t; P, \lambda)} \\ &\quad + e^{\frac{n\lambda^2\sigma_{D,-}^2(t; P, \lambda)}{2}} [Q(\sqrt{n}\lambda\sigma_{D,-}(t; P, \lambda)) - Q(\rho^* + \sqrt{n}\lambda\sigma_{D,-}(t; P, \lambda))] \end{aligned} \quad (3.17)$$

$$\underline{\xi}_{D,-}(t; P, \lambda, n) \triangleq e^{\frac{n\lambda^2\sigma_{D,-}^2(t; P, \lambda)}{2}} Q(\rho_* + \sqrt{n}\lambda\sigma_{D,-}(t; P, \lambda)) \quad (3.18)$$

with  $Q(\rho^*) = \frac{C_{BE}M_{D,-}(t;P,\lambda)}{\sqrt{n}\sigma_{D,-}^3(t;P,\lambda)}$  and  $Q(\rho_*) = \frac{1}{2} - \frac{2C_{BE}M_{D,-}(t;P,\lambda)}{\sqrt{n}\sigma_{D,-}^3(t;P,\lambda)}$ . Similar to the case in Section II, the purpose of introducing above definitions is to utilize the following results, proved as Theorem 8 in [6], which are valid for any  $t \in \mathcal{P}_n$  satisfying (3.6) and (3.15).

(a) *There exists a  $\delta^* > 0$  such that for any  $\delta \in (0, \delta^*]$*

$$r_-(t, \delta) = \frac{1}{2\sigma_D^2(t; P)}\delta^2 + O(\delta^3). \quad (3.19)$$

(b) *For any  $\delta \in (0, \Delta_-^*(t))$ , and any  $x^n \in \mathcal{X}_t^n$ ,*

$$\begin{aligned} \xi_{D,-}(t; P, \lambda, n)e^{-nr_-(t, \delta)} &\geq \Pr \left\{ \frac{1}{n} \ln \frac{p(Y^n|X^n)}{q_t(Y^n)} \leq I(t; P) - \delta \middle| X^n = x^n \right\} \\ &\geq \xi_{D,-}(t; P, \lambda, n)e^{-nr_-(t, \delta)} \end{aligned} \quad (3.20)$$

where  $\lambda = \frac{\partial r_-(t, \delta)}{\partial \delta} > 0$ , and  $Y^n = Y_1 Y_2 \cdots Y_n$  is the output of the DIMC in response to an independent and identically distributed (IID) input  $X^n = X_1 X_2 \cdots X_n$ , the common distribution of each  $X_i$  having  $\mathcal{X}$  as its support set. Moreover, when  $\delta = o(1)$  and  $\delta = \Omega(1/\sqrt{n})$ ,

$$\bar{\xi}_{D,-}(t; P, \lambda, n) = e^{\frac{n\lambda^2\sigma_{D,-}^2(t; P, \lambda)}{2}} Q(\sqrt{n}\lambda\sigma_{D,-}(t; P, \lambda)) (1 + o(1)) \quad (3.21)$$

$$\xi_{D,-}(t; P, \lambda, n) = e^{\frac{n\lambda^2\sigma_{D,-}^2(t; P, \lambda)}{2}} Q(\sqrt{n}\lambda\sigma_{D,-}(t; P, \lambda)) (1 - o(1)) \quad (3.22)$$

and

$$e^{\frac{n\lambda^2\sigma_{D,-}^2(t; P, \lambda)}{2}} Q(\sqrt{n}\lambda\sigma_{D,-}(t; P, \lambda)) = \Theta\left(\frac{1}{\sqrt{n}\lambda}\right) \quad (3.23)$$

with  $\lambda = r'_X(\delta) = \Theta(\delta)$ .

(c) *For any  $\delta \leq c\sqrt{\frac{\ln n}{n}}$ , where  $c < \sigma_D(t; P)$  is a constant, and  $x^n \in \mathcal{X}_t^n$ ,*

$$\begin{aligned} Q\left(\frac{\delta\sqrt{n}}{\sigma_D(t; P)}\right) - \frac{C_{BE}M_D(t; P)}{\sqrt{n}\sigma_D^3(t; P)} &\leq \Pr \left\{ \frac{1}{n} \ln \frac{p(Y^n|X^n)}{q_t(Y^n)} \leq I(t; P) - \delta \middle| X^n = x^n \right\} \\ &\leq Q\left(\frac{\delta\sqrt{n}}{\sigma_D(t; P)}\right) + \frac{C_{BE}M_D(t; P)}{\sqrt{n}\sigma_D^3(t; P)}. \end{aligned} \quad (3.24)$$

Turn our attention to sequences in  $\mathcal{Y}^n$ . For any  $t \in \mathcal{P}_n$  and any  $x^n \in \mathcal{X}_t^n$ , define

$$B_t(x^n, \delta) \triangleq \left\{ y^n : -\infty < \frac{1}{n} \ln \frac{p(y^n|x^n)}{q_t(y^n)} \leq I(t; P) - \delta \right\} \quad (3.25)$$

and

$$\begin{aligned} P_{t, \delta} &\triangleq P_{x^n}(B_t(x^n, \delta)) \\ &= \Pr \left\{ \frac{1}{n} \ln \frac{p(Y^n|X^n)}{q_t(Y^n)} \leq I(t; P) - \delta \middle| X^n = x^n \right\} \end{aligned} \quad (3.26)$$

where  $P_{t,\delta}$  only depends on type  $t$  and  $\delta$ . Since for any  $y^n \in \mathcal{Y}^n$ , the following set

$$\left\{ x^n \in \mathcal{X}_t^n : \frac{1}{n} \ln \frac{p(y^n|x^n)}{q_t(y^n)} \geq I(t; P) - \delta \right\} \quad (3.27)$$

is referred to as a DIMC jar for  $y^n$  based on type  $t$  in [1], we shall call  $B_t(x^n, \delta)$  the *outer mirror image of jar* corresponding to  $x^n$ . Further define

$$B_{t,n,\delta} \triangleq \cup_{x^n \in \mathcal{X}_t^n} B_t(x^n, \delta) \quad (3.28)$$

$$P(B_{t,n,\delta}) \triangleq \int_{y^n \in B_{t,n,\delta}} q_t(y^n) dy^n. \quad (3.29)$$

### B. Converse Coding Theorem

For any channel code  $\mathcal{C}_n$  of block length  $n$  with average word error probability  $P_e(\mathcal{C}_n) = \epsilon_n$ , assume that the message  $M$  is uniformly distributed in  $\{1, 2, \dots, e^{nR(\mathcal{C}_n)}\}$ . Let  $x^n(m)$  be the codeword corresponding to the message  $m$ , and  $\epsilon_{m,n}$  the conditional error probability given message  $m$ . Then

$$\epsilon_n = \mathbf{E}[\epsilon_{M,n}]. \quad (3.30)$$

Let  $\beta_n = \sqrt{\frac{-2 \ln \epsilon_n}{n}}$  and

$$\mathcal{M} \triangleq \{m : \epsilon_{m,n} \leq \epsilon_n(1 + \beta_n)\}. \quad (3.31)$$

Consider a type  $t \in \mathcal{P}_n$  such that

$$|\{m \in \mathcal{M} : t(x^n(m)) = t\}| \geq \frac{|\mathcal{M}|}{(n+1)^{|\mathcal{X}|}}. \quad (3.32)$$

Here and throughout the paper,  $|S|$  denotes the cardinality of a finite set  $S$ . Since  $|\mathcal{P}_n| \leq (n+1)^{|\mathcal{X}|}$ , it follows from the pigeonhole principle that such a type  $t \in \mathcal{P}_n$  exists. In other words, if we classify codewords in  $\{x^n(m) : m \in \mathcal{M}\}$  according to their types, then there is at least one type  $t \in \mathcal{P}_n$  such that the number of codewords in  $\{x^n(m) : m \in \mathcal{M}\}$  with that type is not less than the average.

We are now ready to state our converse theorem for DIMC.

**Theorem 3.** *Given a DIMC, for any channel code  $\mathcal{C}_n$  of block length  $n$  with average word error probability  $P_e(\mathcal{C}_n) = \epsilon_n$ ,*

$$\begin{aligned} R(\mathcal{C}_n) \leq & I(t; P) - \delta - \frac{\ln \epsilon_n - \ln P(B_{t,n,\delta})}{n} + |\mathcal{X}| \frac{\ln(n+1)}{n} \\ & - \frac{\ln \frac{-2 \ln \epsilon_n}{n} - \ln \left(1 + \sqrt{\frac{-2 \ln \epsilon_n}{n}}\right)}{n} \end{aligned} \quad (3.33)$$

for any  $t \in \mathcal{P}_n$  satisfying (3.32), where  $\delta$  is the largest number satisfying

$$\left(1 + 2\sqrt{\frac{-2\ln \epsilon_n}{n}}\right) \epsilon_n \leq P_{t,\delta}. \quad (3.34)$$

Moreover, if a type  $t \in \mathcal{P}_n$  satisfying (3.32) also satisfies (3.6) and (3.15), then the following hold:

1)

$$\begin{aligned} R(\mathcal{C}_n) \leq & I(t; P) - \delta - \frac{\ln \epsilon_n - \ln P(B_{t,n,\delta})}{n} + |\mathcal{X}| \frac{\ln(n+1)}{n} \\ & - \frac{\ln \frac{-2\ln \epsilon_n}{n} - \ln \left(1 + \sqrt{\frac{-2\ln \epsilon_n}{n}}\right)}{n} \end{aligned} \quad (3.35)$$

where  $\delta$  is the solution to

$$\left(1 + 2\sqrt{\frac{-2\ln \epsilon_n}{n}}\right) \epsilon_n = \xi_{D,-}(t; P, \lambda, n) e^{-nr_-(t,\delta)} \quad (3.36)$$

with  $\delta_-(t, \lambda) = \delta$ .

2) When  $\epsilon_n = \frac{e^{-n^\alpha}}{2\sqrt{\pi n^\alpha}} \left(1 - \frac{1}{2n^\alpha}\right)$  for  $\alpha \in (0, 1)$ ,

$$R(\mathcal{C}_n) \leq I(t; P) - \sqrt{2}\sigma_D(t; P)n^{-\frac{1-\alpha}{2}} + O(n^{-(1-\alpha)}). \quad (3.37)$$

3) When  $\epsilon_n = \frac{n^{-\alpha}}{2\sqrt{\pi \alpha \ln n}} \left(1 - \frac{1}{2\alpha \ln n}\right)$  for  $\alpha > 0$ ,

$$R(\mathcal{C}_n) \leq I(t; P) - \sigma_D(t; P) \sqrt{\frac{2\alpha \ln n}{n}} + O\left(\frac{\ln n}{n}\right). \quad (3.38)$$

4) When  $\epsilon_n = \epsilon$  satisfying  $\epsilon + \frac{1}{\sqrt{n}} \left(2\epsilon\sqrt{-2\ln \epsilon} + \frac{C_{BE}M_D(t; P)}{\sigma_D^3(t; P)}\right) < 1$ ,

$$\begin{aligned} R(\mathcal{C}_n) \leq & I(t; P) - \frac{\sigma_D(t; P)}{\sqrt{n}} Q^{-1} \left( \epsilon + \frac{1}{\sqrt{n}} \left( 2\epsilon\sqrt{-2\ln \epsilon} + \frac{C_{BE}M_D(t; P)}{\sigma_D^3(t; P)} \right) \right) \\ & + (|\mathcal{X}| + 1) \frac{\ln n}{n} - \frac{\ln \epsilon}{n} \end{aligned} \quad (3.39)$$

$$= I(t; P) - \frac{\sigma_D(t; P)}{\sqrt{n}} Q^{-1}(\epsilon) + (|\mathcal{X}| + 1) \frac{\ln n}{n} + O(n^{-1}). \quad (3.40)$$

*Proof:* We again apply the outer mirror image of jar converse-proof technique. By Markov inequality,

$$\Pr\{M \in \mathcal{M}\} \geq \frac{\beta_n}{1 + \beta_n} \text{ and } |\mathcal{M}| \geq e^{nR(\mathcal{C}_n) + \ln \frac{\beta_n}{1 + \beta_n}}. \quad (3.41)$$

For any  $t \in \mathcal{P}_n$  satisfying (3.32), let

$$\mathcal{M}_t \triangleq \{m : \epsilon_{m,n} \leq \epsilon_n(1 + \beta_n), t(x^n(m)) = t\}. \quad (3.42)$$

Then

$$|\mathcal{M}_t| \geq \frac{|\mathcal{M}|}{(n+1)^{|\mathcal{X}|}} \geq e^{nR(\mathcal{C}_n) + \ln \frac{\beta_n}{1+\beta_n} - |\mathcal{X}| \ln(n+1)}. \quad (3.43)$$

Denote the decision region for message  $m \in \mathcal{M}_t$  as  $D_m$ . Now for any  $m \in \mathcal{M}_t$ ,

$$\begin{aligned} P_{x^n(m)}(B_t(x^n(m), \delta) \cap D_m) &= P_{x^n(m)}(B_t(x^n(m), \delta)) - P_{x^n(m)}(B_t(x^n(m), \delta) \cap D_m^c) \\ &\geq P_{x^n(m)}(B_t(x^n(m), \delta)) - \epsilon_{m,n} \\ &\geq P_{x^n(m)}(B_t(x^n(m), \delta)) - \epsilon_n(1 + \beta_n) \end{aligned} \quad (3.44)$$

At this point, we select  $\delta$  such that for any  $x^n \in \mathcal{X}_t^n$ ,

$$P_{x^n}(B_t(x^n, \delta)) = P_{t,\delta} \geq \epsilon_n(1 + 2\beta_n). \quad (3.45)$$

Substituting (3.45) into (3.44), we have

$$P_{x^n(m)}(B_t(x^n(m), \delta) \cap D_m) \geq \beta_n \epsilon_n. \quad (3.46)$$

By the fact that  $D_m$  are disjoint for different  $m$  and

$$\cup_{m \in \mathcal{M}_t} (D_m \cap B_t(x^n(m), \delta)) \subseteq B_{t,n,\delta}, \quad (3.47)$$

we have

$$\begin{aligned} P(B_{t,n,\delta}) &= \int_{B_{t,n,\delta}} q_t(y^n) dy^n \\ &\geq \sum_{m \in \mathcal{M}_t} \int_{B(x^n(m), \delta) \cap D_m} q_t(y^n) dy^n \\ &\geq \sum_{m \in \mathcal{M}_t} \int_{B(x^n(m), \delta) \cap D_m} p(y^n | x^n(m)) e^{-n(I(t;P) - \delta)} dy^n \\ &= \sum_{m \in \mathcal{M}_t} e^{-n(I(t;P) - \delta)} \int_{B(x^n(m), \delta) \cap D_m} p(y^n | x^n(m)) dy^n \\ &= \sum_{m \in \mathcal{M}_t} e^{-n(I(t;P) - \delta)} P_{x^n(m)}(B(x^n(m), \delta) \cap D_m) \\ &\geq \sum_{m \in \mathcal{M}_t} e^{-n(I(t;P) - \delta)} \beta_n \epsilon_n = |\mathcal{M}_t| e^{-n(I(t;P) - \delta)} \beta_n \epsilon_n \end{aligned} \quad (3.48)$$

which implies that

$$|\mathcal{M}_t| \leq e^{n(I(t;P) - \delta) - \ln \beta_n - \ln \epsilon_n + \ln P(B_{t,n,\delta})}. \quad (3.49)$$



Then combining (3.43) and (3.49) yields

$$R(\mathcal{C}_n) \leq I(t; P) - \delta - \frac{\ln \epsilon_n - \ln P(B_{t,n,\delta})}{n} - \frac{\ln \frac{\beta_n}{1+\beta_n}}{n} - \frac{\ln \beta_n}{n} + |\mathcal{X}| \frac{\ln(n+1)}{n}. \quad (3.50)$$

Since  $\beta_n = \sqrt{\frac{-2 \ln \epsilon_n}{n}}$  by definition, (3.33) and (3.34) directly come from (3.50) and (3.45).

- 1) According to (3.20), it can be seen that selecting  $\delta$  to be the solution to (3.36) will suffice (3.45). Consequently, (3.35) is proved.
- 2) The proof is essentially the same as that for part 2) of Theorem 1, where we can show that

$$P_{t,\delta} \geq \left(1 + 2\sqrt{\frac{-2 \ln \epsilon_n}{n}}\right) \epsilon_n \quad (3.51)$$

when  $\epsilon_n = \frac{e^{-n^\alpha}}{2\sqrt{\pi n^\alpha}} \left(1 - \frac{1}{2n^\alpha}\right)$  and  $\delta = \sqrt{2}\sigma_D(t; P)n^{-\frac{1-\alpha}{2}} - \eta n^{-(1-\alpha)}$  for some constant  $\eta$ .

- 3) Apply the trivial bound  $P(B_{t,n,\delta}) \leq 1$ . Then similar to the proof for part 3) of Theorem 1, one can verify that by making  $\delta = \sigma_{D,-}(t; P)\sqrt{\frac{2\alpha \ln n}{n}} - \eta \frac{\ln n}{n}$  for some properly chosen constant  $\eta$ ,

$$\begin{aligned} P_{t,\delta} &\geq \xi_{D,-} \left(t; P, \frac{\partial r_-(t, \delta)}{\partial \delta}, n\right) e^{-nr_-(t, \delta)} \\ &\geq \left(1 + 2\sqrt{\frac{-2 \ln \epsilon_n}{n}}\right) \epsilon \end{aligned} \quad (3.52)$$

for  $\epsilon_n = \frac{n^{-\alpha}}{2\sqrt{\pi \alpha \ln n}} \left(1 - \frac{1}{2\alpha \ln n}\right)$ , where (3.19), (3.22) and (3.23) are utilized.

- 4) According to (3.45), we should select  $\delta$  such that

$$P_{t,\delta} \geq \left(1 + 2\sqrt{\frac{-2 \ln \epsilon}{n}}\right) \epsilon. \quad (3.53)$$

Now by (3.24),

$$\delta = \frac{\sigma_D(t; P)}{\sqrt{n}} Q^{-1} \left( \epsilon + \frac{1}{\sqrt{n}} \left( 2\epsilon \sqrt{-2 \ln \epsilon} + \frac{C_{BE} M_D(t; P)}{\sigma_D^3(t; P)} \right) \right) \quad (3.54)$$

will guarantee (3.53). Consequently, (3.39) is proved by substituting (3.53) and  $\epsilon_n = \epsilon$  into (3.50) and applying the trivial bound  $P(B_{t,n,\delta}) \leq 1$ , and (3.40) is yielded by the property of  $Q^{-1}$  function shown in the proof of Theorem 1. ■

**Remark 10.** Remarks similar to Remarks 2 and 3 can be drawn here too for Theorem 3.

For maximal error probability, we have the following corollary, which can be proved similarly.

**Corollary 2.** *Given a DIMC, for any channel code  $\mathcal{C}_n$  of block length  $n$  with maximum error probability  $P_m(\mathcal{C}_n) = \epsilon_n$ ,*

$$R(\mathcal{C}_n) \leq I(t; P) - \delta - \frac{\ln \epsilon_n - \ln P(B_{t,n,\delta})}{n} + |\mathcal{X}| \frac{\ln(n+1)}{n} - \frac{\ln \sqrt{\frac{-2 \ln \epsilon_n}{n}}}{n} \quad (3.55)$$

for any  $t \in \mathcal{P}_n$  such that there are at least  $(n+1)^{-|\mathcal{X}|}$  portion of codewords in  $\mathcal{C}_n$  with type  $t$ , where  $\delta$  is the largest number satisfying

$$\left(1 + \sqrt{\frac{-2 \ln \epsilon_n}{n}}\right) \epsilon_n \leq P_{t,\delta}. \quad (3.56)$$

Moreover, if  $t \in \mathcal{P}_n$  satisfies (3.6) and (3.15), then the following hold:

1)

$$R(\mathcal{C}_n) \leq I(t; P) - \delta - \frac{\ln \epsilon_n - \ln P(B_{t,n,\delta})}{n} + |\mathcal{X}| \frac{\ln(n+1)}{n} - \frac{\ln \sqrt{\frac{-2 \ln \epsilon_n}{n}}}{n} \quad (3.57)$$

where  $\delta$  is the solution to

$$\left(1 + \sqrt{\frac{-2 \ln \epsilon_n}{n}}\right) \epsilon_n = \xi_{D,-}(t; P, \lambda, n) e^{-nr_-(t,\delta)} \quad (3.58)$$

with  $\delta_-(t, \lambda) = \delta$ .

2) When  $\epsilon_n = \epsilon$  satisfying  $\epsilon + \frac{1}{\sqrt{n}} \left( \epsilon \sqrt{-2 \ln \epsilon} + \frac{C_{BE} M_D(t; P)}{\sigma_D^3(t; P)} \right) < 1$ ,

$$\begin{aligned} R(\mathcal{C}_n) &\leq I(t; P) - \frac{\sigma_D(t; P)}{\sqrt{n}} Q^{-1} \left( \epsilon + \frac{1}{\sqrt{n}} \left( \epsilon \sqrt{-2 \ln \epsilon} + \frac{C_{BE} M_D(t; P)}{\sigma_D^3(t; P)} \right) \right) \\ &\quad + (|\mathcal{X}| + 0.5) \frac{\ln n}{n} - \frac{\ln \epsilon}{n} \end{aligned} \quad (3.59)$$

$$= I(t; P) - \frac{\sigma_D(t; P)}{\sqrt{n}} Q^{-1}(\epsilon) + (|\mathcal{X}| + 0.5) \frac{\ln(n+1)}{n} + O(n^{-1}). \quad (3.60)$$

### C. Taylor-Type Expansion

Fix a DIMC  $P = \{p(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$  with its capacity  $C_{\text{DIMC}} > 0$ . For any block length  $n$  and average error probability  $\epsilon$ , let  $R_n(\epsilon)$  be the best coding rate achievable with block length  $n$  and average error probability  $\leq \epsilon$ , as defined in (2.66). In this subsection, we extend Theorem 2 to establish a Taylor-type expansion of  $R_n(\epsilon)$  in the case of DIMC.

We begin with reviewing the non-asymptotic achievability of jar decoding established in [1]. It has been proved in [1] that under jar decoding, Shannon random codes  $\mathcal{C}_n$  of block length  $n$  based on any type  $t \in \mathcal{P}_n$  satisfying (3.6) and (3.15) have the following performance:

1)

$$R(\mathcal{C}_n) \geq I(t; P) - \delta - r_-(t, \delta) - \frac{(0.5 + |\mathcal{X}|) \ln(n+1) - \ln \frac{2(1-C_{BE})M_{D,-}(t; P, \lambda)}{\sqrt{n}\sigma_{D,-}^3(t; P, \lambda)}}{n} \quad (3.61)$$

while maintaining

$$P_e(\mathcal{C}_n) \leq \left( \bar{\xi}_{D,-}(t; P, \lambda, n) + \frac{2(1-C_{BE})M_{D,-}(t; P, \lambda)}{\sqrt{n}\sigma_{D,-}^3(t; P, \lambda)} \right) e^{-nr_-(t, \delta)} \quad (3.62)$$

for any  $\delta \in (0, \Delta_-^*(t))$ , where  $\lambda = \frac{\partial r_-(t, \delta)}{\partial \delta}$  satisfying  $\delta_-(t, \lambda) = \delta$ .

2)

$$R(\mathcal{C}_n) \geq I(t; P) - \sigma_D(t; P) \sqrt{\frac{2\alpha \ln n}{n}} - \frac{(0.5 + \alpha + |\mathcal{X}|) \ln(n+1)}{n} - O\left(\frac{\ln \ln n}{n}\right) \quad (3.63)$$

while maintaining

$$P_e(\mathcal{C}_n) \leq \frac{n^{-\alpha}}{2\sqrt{\pi\alpha \ln n}} + O\left(n^{-\alpha} \frac{\ln n}{\sqrt{n}}\right) = \Theta\left(\frac{n^{-\alpha}}{\sqrt{\ln n}}\right) \quad (3.64)$$

for any  $\alpha \geq 0$ .

3)

$$R(\mathcal{C}_n) \geq I(t; P) - \frac{c}{\sqrt{n}} - \left(\frac{1}{2} + |\mathcal{X}|\right) \frac{\ln(n+1)}{n} - \frac{1}{n} \ln \frac{(1-C_{BE})M_D(t; P)}{\sigma_D^3(t; P)} \quad (3.65)$$

while maintaining

$$P_e(\mathcal{C}_n) \leq Q\left(\frac{c}{\sigma_D(t; P)}\right) + \frac{M_D(t; P)}{\sigma_D^3(t; P)} \frac{1}{\sqrt{n}} \quad (3.66)$$

for any real number  $c$ .

By combining (3.61) and (3.62) with (3.33) and (3.34) or with (3.35) and (3.36), it is expected that  $R_n(\epsilon)$  would be expanded as

$$R_n(\epsilon) = I(t; P) - \delta + o(\delta) \quad (3.67)$$

for some  $t \in \mathcal{P}$ , where  $\delta$  is defined according to (3.62), (3.34), or (3.36). In the rest of this subsection, we shall demonstrate with mathematic rigor that this is indeed the case. To simplify our argument, we impose the following conditions\* on the channel:

(C1) For any  $t \in \mathcal{P}$ ,  $M_D(t; P) < \infty$ .

\*Some of these conditions, for example, Condition C3, can be relaxed. Here we choose not to do so in order not to make our subsequent argument unnecessary complicated.

- (C2)  $\sigma_D^2(t; P) = 0$  implies  $I(t; P) = 0$ .
- (C3) For any  $t \in \mathcal{P}$ ,  $\lambda_-^*(t; P) = +\infty$ .
- (C4) There exists  $\lambda^* > 0$  such that  $\delta_-(t, \lambda)$ ,  $\sigma_{D,-}^2(t; P, \lambda)$ ,  $M_{D,-}(t; P, \lambda)$ ,  $\hat{M}_{D,-}(t; P, \lambda)$ , and  $r_-(t, \delta_-(t, \lambda))$  are continuous functions of  $t$  and  $\lambda$  over  $(t, \lambda) \in \mathcal{P} \times [0, \lambda^*]$ .
- (C5) There exists  $s^* > 0$  such that  $r_-^{-1}(t, s)$  is a continuous function of  $t$  and  $s$  over  $(t, s) \in \mathcal{P} \times [0, s^*]$ , where  $r_-^{-1}(t, \cdot)$  is an inverse function of  $r_-(t, \cdot)$ .

Since  $r_-(t, \delta)$  is a continuous and strictly increasing function of  $\delta$  before it reaches  $+\infty$ —which may or may not happen—it can be easily verified that for any  $s \geq 0$

$$\begin{aligned} r_-^{-1}(t, s) &= \max\{\delta : r_-(t, \delta) \leq s\} \\ &= \inf\{\delta : r_-(t, \delta) > s\}. \end{aligned} \quad (3.68)$$

In view of the definitions and properties of  $\delta_-(t, \lambda)$ ,  $\sigma_{D,-}^2(t; P, \lambda)$ ,  $M_{D,-}(t; P, \lambda)$ ,  $\hat{M}_{D,-}(t; P, \lambda)$ , and  $r_-(t, \delta)$  (see [6] for details and examples), Conditions (C1) to (C5) are generally met by most channels, particularly by channels with discrete output alphabets, and discrete input additive white Gaussian channels.

To characterize  $\delta$  in (3.67) analytically, we need a counterpart of Lemma 1. To this end, define for any  $0 < c < C_{\text{DIMC}}$

$$\mathcal{P}(c) \triangleq \{t \in \mathcal{P} : I(t; P) \geq c\} \quad (3.69)$$

$$\mathcal{P}_n(c) \triangleq \{t \in \mathcal{P}_n : I(t; P) \geq c\} \quad (3.70)$$

and for any type  $t \in \mathcal{P}$  satisfying  $\sigma_D^2(t; P) > 0$

$$g_{t;P,n}(\delta) \triangleq e^{\frac{n\lambda^2\sigma_{D,-}^2(t;P,\lambda)}{2}} Q(\sqrt{n}\lambda\sigma_{D,-}(t;P,\lambda))e^{-nr_-(t,\delta)} \quad (3.71)$$

where  $\lambda = \frac{\partial r_-(t,\delta)}{\partial \delta}$ . Note that  $\mathcal{P}(c)$  is a closed set, and it follows from Condition (C2) that  $\sigma_D^2(t; P) > 0$  for any  $t \in \mathcal{P}(c)$ . Interpret  $g_{t;P,n}(\delta)$  as a function of  $\lambda$  through  $\delta = \delta_-(t, \lambda)$ . Then we have the following lemma.

**Lemma 2.** *There exists  $\lambda^+ > 0$  such that for any  $n > 0$  and  $t \in \mathcal{P}(c)$ ,  $g_{t;P,n}(\delta_-(t, \lambda))$  is a strictly decreasing function of  $\lambda$  over  $\lambda \in [0, \lambda^+]$ .*

*Proof:* The proof is in parallel with that of Lemma 1. As such, we point out only places where differences occur. In the place of (2.80), we now have

$$\frac{d\sigma_{D,-}^2(t; P, \lambda)}{d\lambda} = -\hat{M}_{D,-}(t; P, \lambda). \quad (3.72)$$

In parallel with (2.81) and (2.82), we now have for any  $t \in \mathcal{P}(c)$

$$\begin{aligned} & \frac{dg_{t;P,n}(\delta_-(t, \lambda))}{d\lambda} \\ & \leq e^{-nr_-(t, \delta_-(t, \lambda))} \frac{\sqrt{n}\sigma_{D,-}(t; P, \lambda)}{\sqrt{2\pi}} \left( \left| -\frac{\lambda \frac{d\sigma_{D,-}^2(t; P, \lambda)}{d\lambda}}{2\sigma_{D,-}^2(t; P, \lambda) (1 + n\lambda^2\sigma_{D,-}^2(t; P, \lambda))} \right| - 1 \right) \\ & = e^{-nr_-(t, \delta_-(t, \lambda))} \frac{\sqrt{n}\sigma_{D,-}(t; P, \lambda)}{\sqrt{2\pi}} \left( \left| \frac{\lambda \hat{M}_{D,-}(t; P, \lambda)}{2\sigma_{D,-}^2(t; P, \lambda) (1 + n\lambda^2\sigma_{D,-}^2(t; P, \lambda))} \right| - 1 \right) \quad (3.73) \end{aligned}$$

$$\leq e^{-nr_-(t, \delta_-(t, \lambda))} \frac{\sqrt{n}\sigma_{D,-}(t; P, \lambda)}{\sqrt{2\pi}} \left( \left| \frac{\lambda \hat{M}_{D,-}(t; P, \lambda)}{2\sigma_{D,-}^2(t; P, \lambda)} \right| - 1 \right). \quad (3.74)$$

Since  $\mathcal{P}(c)$  is closed, it then follows from Condition (C4) that there is a  $\lambda^+ > 0$  such that for any  $\lambda \in [0, \lambda^+]$  and any  $t \in \mathcal{P}(c)$

$$\left| \frac{\lambda \hat{M}_{D,-}(t; P, \lambda)}{2\sigma_{D,-}^2(t; P, \lambda)} \right| - 1 < 0$$

and hence

$$\frac{dg_{t;P,n}(\delta_-(t, \lambda))}{d\lambda} < 0$$

for any  $n > 0$ . This completes the proof of Lemma 2.  $\blacksquare$

**Remark 11.** In view of (3.73), it is clear that when  $n$  is large,  $g_{t;P,n}(\delta_-(t, \lambda))$  is a strictly decreasing function of  $\lambda$  over an interval even larger than  $[0, \lambda^+]$  for each and every  $t \in \mathcal{P}(c)$ .

Now let

$$\epsilon_n^+ \triangleq \max\{g_{t;P,n}(\delta_-(t, \lambda^+/2)) : t \in \mathcal{P}(c)\}$$

which, in view of Condition (C4) and the fact that  $\mathcal{P}(c)$  is closed, is well defined and also an exponential function of  $n$ . For any  $\epsilon_n^+ \leq \epsilon \leq 1/2$  and  $t \in \mathcal{P}(c)$ , let  $\delta_{t,n}(\epsilon)$  be the unique solution to

$$g_{t;P,n}(\delta) = \epsilon. \quad (3.75)$$

Further define

$$s(c) \triangleq \max \left\{ s : 0 < s \leq s^*, r_-^{-1}(t, s) \leq \frac{C_{\text{DIMC}} - c}{2} \quad \forall t \in \mathcal{P} \right\} \quad (3.76)$$

and let  $\epsilon_n(c)$  be the unique solution  $\epsilon$  to

$$\frac{-\ln \epsilon \left( 1 + 2\sqrt{\frac{-2\ln \epsilon}{n}} \right)}{n} = s(c). \quad (3.77)$$

It is easy to see that in view of Condition (C5),  $s(c) > 0$  is well defined and once again  $\epsilon_n(c)$  is also an exponential function of  $n$ . Let  $\epsilon_n^u < 1$  be the unique solution  $\epsilon$  to

$$\epsilon \left( 1 + 2\sqrt{\frac{-2\ln \epsilon}{n}} \right) = 1. \quad (3.78)$$

Note that

$$\max\{I(t; P) : t \in \mathcal{P}_n\} = C_{\text{DIMC}} - O\left(\frac{1}{n^2}\right).$$

Let  $N(c)$  be the smallest integer  $N > 0$  such that

$$\max\{I(t; P) : t \in \mathcal{P}_n\} \geq C_{\text{DIMC}} - \frac{C_{\text{DIMC}} - c}{2} \quad (3.79)$$

for all  $n \geq N$ . Then we have the following Taylor-type expansion of  $R_n(\epsilon)$ .

**Theorem 4.** *For any  $n \geq N(c)$  and any  $\max\{\epsilon_n^+, \epsilon_n(c)\} \leq \epsilon < \epsilon_n^u$ , let*

$$t^* \triangleq \arg \max_{t \in \mathcal{P}_n(c)} [I(t; P) - \delta_{t,n}(\epsilon)] \quad (3.80)$$

$$t^\# \triangleq \arg \max_{t \in \mathcal{P}_n(c)} \left[ I(t; P) - \frac{\sigma_D(t; P)}{\sqrt{n}} Q^{-1}(\epsilon) \right]. \quad (3.81)$$

Then

$$|R_n(\epsilon) - (I(t^*; P) - \delta_{t^*,n}(\epsilon))| \leq o(\delta_{t^*,n}(\epsilon)) \quad (3.82)$$

where

$$o(\delta_{t^*,n}(\epsilon)) = r_-(t^*, \delta_{t^*,n}(\epsilon)) + \frac{(|\mathcal{X}| + 1.5) \ln(n+1) + d_1}{n} \quad (3.83)$$

if  $\epsilon \leq \frac{1}{3}$ , and

$$\left| R_n(\epsilon) - \left( I(t^\#; P) - \frac{\sigma_D(t^\#; P)}{\sqrt{n}} Q^{-1}(\epsilon) \right) \right| \leq \frac{(|\mathcal{X}| + 1) \ln(n+1) + d_2}{n} \quad (3.84)$$

otherwise, where  $d_1$  and  $d_2$  are constants depending on the channel, but independent of  $n$  and  $\epsilon$ .

*Proof:* For any  $t \in \mathcal{P}_n$  and  $0 < \epsilon < 1$ , let

$$\delta_{t,n}^P(\epsilon) = \sup \left\{ \delta > 0 : P_{t,\delta} \geq \left( 1 + 2\sqrt{\frac{-2\ln \epsilon}{n}} \right) \epsilon \right\}.$$

By Theorem 3 and the trivial bound  $P(B_{t,n,\delta}) \leq 1$ , it is not hard to verify that

$$R_n(\epsilon) \leq \max_{t \in \mathcal{P}_n} [I(t; P) - \delta_{t,n}^P] - \frac{\ln \epsilon + \ln \frac{-2\ln \epsilon}{n}}{n} + \frac{\ln \left( 1 + \sqrt{\frac{-2\ln \epsilon}{n}} \right) + |\mathcal{X}| \ln(n+1)}{n}. \quad (3.85)$$

Let us now examine

$$\max_{t \in \mathcal{P}_n} [I(t; P) - \delta_{t,n}^P].$$

In view of the Chernoff bound (see Theorem 8 in [6]),

$$P_{t,\delta} \leq e^{-nr_-(t,\delta)}$$

for any  $t \in \mathcal{P}_n$  and  $\delta > 0$ , which, together with (3.68), implies

$$\delta_{t,n}^P \leq r_-^{-1} \left( t, \frac{-\ln \left( 1 + 2\sqrt{\frac{-2\ln \epsilon}{n}} \right) \epsilon}{n} \right) \quad (3.86)$$

$$\leq r_-^{-1}(t, s(c)) \quad (3.87)$$

$$\leq \frac{C_{\text{DIMC}} - c}{2} \quad (3.88)$$

whenever  $\max\{\epsilon_n^+, \epsilon_n(c)\} \leq \epsilon < \epsilon_n^u$ . In the above derivation, (3.86) is due to (3.68); and (3.87) and (3.88) follow from (3.76), (3.77), and (3.78). Therefore,

$$\begin{aligned} \max_{t \in \mathcal{P}_n} [I(t; P) - \delta_{t,n}^P] &\geq \max_{t \in \mathcal{P}_n} I(t; P) - \frac{C_{\text{DIMC}} - c}{2} \\ &\geq c \end{aligned} \quad (3.89)$$

where the last inequality is due to (3.79). In view of (3.89), it is not hard to see that for any  $t \in \mathcal{P}_n$  achieving  $\max_{t \in \mathcal{P}_n} [I(t; P) - \delta_{t,n}^P]$ ,

$$I(t; P) \geq c + \delta_{t,n}^P \geq c$$

and hence

$$\max_{t \in \mathcal{P}_n} [I(t; P) - \delta_{t,n}^P] = \max_{t \in \mathcal{P}_n(c)} [I(t; P) - \delta_{t,n}^P]$$

which, together with (3.85), implies

$$R_n(\epsilon) \leq \max_{t \in \mathcal{P}_n(c)} [I(t; P) - \delta_{t,n}^P] - \frac{\ln \epsilon + \ln \frac{-2\ln \epsilon}{n}}{n} + \frac{\ln \left( 1 + \sqrt{\frac{-2\ln \epsilon}{n}} \right) + |\mathcal{X}| \ln(n+1)}{n}. \quad (3.90)$$

When  $\epsilon > \frac{1}{3}$ , it follows from (3.24) and (3.54) that for any  $t \in \mathcal{P}_n(c)$ ,

$$\begin{aligned} \delta_{t,n}^P &\geq \frac{\sigma_D(t; P)}{\sqrt{n}} Q^{-1} \left( \epsilon + \frac{1}{\sqrt{n}} \left( 2\epsilon \sqrt{-2\ln \epsilon} + \frac{C_{BE} M_D(t; P)}{\sigma_D^3(t; P)} \right) \right) \\ &\geq \frac{\sigma_D(t; P)}{\sqrt{n}} Q^{-1}(\epsilon) - \sqrt{2\pi} e^{\frac{[Q^{-1}(\epsilon)]^2}{2}} \frac{\sigma_D(t; P)}{n} \left( 2\epsilon \sqrt{-2\ln \epsilon} + \frac{C_{BE} M_D(t; P)}{\sigma_D^3(t; P)} \right) \end{aligned} \quad (3.91)$$

Since  $\mathcal{P}(c)$  is closed, it follows Condition (C4) that  $\sigma_D(t; P)$  and  $\frac{M_D(t; P)}{\sigma_D^3(t; P)}$  are bounded over  $\mathcal{P}(c)$ . Plugging (3.91) into (3.90) yields

$$R_n(\epsilon) \leq \max_{t \in \mathcal{P}_n(c)} \left[ I(t; P) - \frac{\sigma_D(t; P)}{\sqrt{n}} Q^{-1}(\epsilon) \right] + \frac{(|\mathcal{X}| + 1) \ln(n + 1) + d}{n}$$

for some constant  $d$ , which, together with the achievability in (3.65) and (3.66), implies (3.84).

Now let us focus on the case when  $\epsilon \leq \frac{1}{3}$ . For any  $t \in \mathcal{P}(c)$ , let  $\underline{\delta}_{t,n}(\epsilon)$  be the unique solution to

$$\left( 1 + 2\sqrt{\frac{-2 \ln \epsilon}{n}} \right) \epsilon = \xi_{D,-}(t; P, \lambda, n) e^{-nr_{-(t,\delta)}} \quad (3.92)$$

where  $\lambda = \frac{\partial r_{-(t,\delta)}}{\partial \delta}$ . By following the argument in the proof of Theorem 2, it is not hard to verify that for any  $t \in \mathcal{P}_n(c)$

$$\delta_{t,n}^P(\epsilon) \geq \underline{\delta}_{t,n}(\epsilon) \geq \delta_{t,n}(\epsilon) - \frac{d}{n} \quad (3.93)$$

for some constant  $d$  independent of  $n$ ,  $\epsilon$ , and  $t$ . Plugging (3.93) into (3.90) then yields

$$R_n(\epsilon) \leq I(t^*; P) - \delta_{t^*,n}(\epsilon) - \frac{\ln \epsilon + \ln \frac{-2 \ln \epsilon}{n}}{n} + \frac{\sqrt{\frac{-2 \ln \epsilon}{n}} + |\mathcal{X}| \ln(n + 1) + d}{n}. \quad (3.94)$$

In the meantime,

$$\begin{aligned} \epsilon &= g_{t^*; P, n}(\delta_{t^*, n}) \\ &\geq \frac{1}{\sqrt{2\pi} \left( \sqrt{n} \lambda_{t^*, n} \sigma_{D,-}(t^*; P, \lambda_{t^*, n}) + \frac{1}{\sqrt{n} \lambda_{t^*, n} \sigma_{D,-}(t^*; P, \lambda_{t^*, n})} \right)} e^{-nr_{-(t^*, \delta_{t^*, n}(\epsilon))}} \end{aligned} \quad (3.95)$$

where  $\lambda_{t^*, n} = \frac{\partial r_{-(t^*, \delta)}}{\partial \delta} \Big|_{\delta = \delta_{t^*, n}(\epsilon)}$ . Consequently,

$$\begin{aligned} \frac{-\ln \epsilon}{n} &\leq r_{-(t^*, \delta_{t^*, n}(\epsilon))} + \frac{\ln \left[ \sqrt{2\pi} \left( \sqrt{n} \lambda_{t^*, n} \sigma_{D,-}(t^*; P, \lambda_{t^*, n}) + \frac{1}{\sqrt{n} \lambda_{t^*, n} \sigma_{D,-}(t^*; P, \lambda_{t^*, n})} \right) \right]}{n} \\ &\leq r_{-(t^*, \delta_{t^*, n}(\epsilon))} + \frac{\ln n}{2n} + \frac{\eta_1}{n} \end{aligned} \quad (3.96)$$

where  $\eta_1$  is a constant independent of  $n$ ,  $\epsilon$ , and  $t^*$ . Now substituting (3.96) and  $\epsilon \leq \frac{1}{3}$  into (3.94) yields

$$\begin{aligned} R_n(\epsilon) &\leq I(t^*; P) - \delta_{t^*, n}(\epsilon) + r_{-(t^*, \delta_{t^*, n}(\epsilon))} \\ &\quad + \frac{-\ln \frac{2 \ln 3}{n} + \eta_1 + \sqrt{r_{-(t^*, \delta_{t^*, n}(\epsilon))} + \frac{1}{2e}} + \frac{\eta_1}{n} + \frac{1}{2} \ln n + |\mathcal{X}| \ln(n + 1) + d}{n} \\ &\leq I(t^*; P) - \delta_{t^*, n}(\epsilon) + r_{-(t^*, \delta_{t^*, n}(\epsilon))} + \frac{d_1 + (|\mathcal{X}| + \frac{3}{2}) \ln(n + 1)}{n} \end{aligned} \quad (3.97)$$



for some constant  $\bar{d}_1$  independent of  $n$ ,  $\epsilon$ , and  $t^*$ , where the last inequality is due to the fact that in view of Condition (4),  $r_-(t^*, \delta_{t^*,n}(\epsilon))$  is bounded over  $t \in \mathcal{P}(c)$  and  $\epsilon \geq \max\{\epsilon_n^+, \epsilon_n(c)\}$ .

To complete the proof, let us go back to the achievability given in (3.61) and (3.62). Now choose  $t$  to be  $t^*$ , and follow the argument in the proof of Theorem 2. Then it is not hard to show that

$$R_n(\epsilon) \geq I(t^*; P) - \delta_{t^*,n}(\epsilon) - r_-(t^*, \delta_n(\epsilon)) - \frac{(|\mathcal{X}| + 1) \ln(n + 1) + \bar{d}_1}{n} \quad (3.98)$$

where  $\bar{d}_1$  is a constant independent of  $n$ ,  $\epsilon$ , and  $t^*$ . Combining (3.98) with (3.97) completes the proof of Theorem 4.  $\blacksquare$

Remarks similar to those immediately after Theorem 2 also apply here. In particular, Theorem 4 and the achievability of jar decoding given in (3.61) and (3.62) to (3.65) and (3.66) once again imply that jar decoding is indeed optimal up to the second order coding performance in the non-asymptotical regime for any DIMC. In addition, the following remarks are helpful to the computation of the Taylor-type expansion of  $R_n(\epsilon)$  as expressed in (3.80) to (3.84).

**Remark 12.** When  $I(t; P)$ ,  $\delta_-(t, \lambda)$ ,  $\sigma_{D,-}^2(t; P, \lambda)$ ,  $M_{D,-}(t; P, \lambda)$ ,  $\hat{M}_{D,-}(t; P, \lambda)$ , and  $r_-(t, \delta_-(t, \lambda))$  are all continuously differentiable with respect to  $t$  over  $t \in \mathcal{P}(c)$  and  $\lambda \in [0, \lambda^*]$ , which is true for most channels including particularly channels with discrete output alphabets, and discrete input additive white Gaussian channels,  $\mathcal{P}_n(c)$  in the definitions of  $t^*$  and  $t^\#$  can be replaced by  $\mathcal{P}(c)$ . Thus, in this case,

$$t^* \triangleq \arg \max_{t \in \mathcal{P}(c)} [I(t; P) - \delta_{t,n}(\epsilon)] \quad (3.99)$$

$$t^\# \triangleq \arg \max_{t \in \mathcal{P}(c)} \left[ I(t; P) - \frac{\sigma_D(t; P)}{\sqrt{n}} Q^{-1}(\epsilon) \right]. \quad (3.100)$$

Hereafter, we shall assume that the channel satisfies this continuously differentiable condition, and use (3.99) and (3.80), or (3.100) and (3.81) interchangeably.

**Remark 13.** It is worth pointing out the impact of  $c$  on the maximization problems given in (3.99), (3.80), (3.100), and (3.81). In view of the definitions of  $s(c)$  and  $\epsilon_n(c)$  in (3.76) and (3.77), it is not hard to see that when  $\epsilon$  is relatively large with respect to  $n$  (in the sense that  $\frac{-\ln \epsilon}{n}$  is small), one can select  $c$  to be close to  $C_{\text{DIMC}}$ . In this case, it suffices to search a small range  $\mathcal{P}(c)$  for optimal  $t^*$ . On the other hand, when  $\epsilon$  is relatively small with respect to  $n$ , e.g.,

a exponential function of  $n$ ,  $c$  should be selected to be far below  $C_{\text{DIMC}}$  and hence one has to search a large range  $\mathcal{P}(c)$  for optimal  $t^*$ .

**Remark 14.** When the Taylor-type expansion of  $R_n(\epsilon)$  in Theorem 4 is applied to the case of BIMSC, it yields essentially the same result as in Theorem 2, with explanation as follows. For any BIMSC,  $t(0)$  fully characterizes the type  $t$ . Then by symmetry,  $\frac{\partial \delta_{t,n}(\epsilon)}{\partial t(0)} = 0$  at  $t(0) = 0.5$  for any  $n$  and  $\epsilon$ . Note that  $\delta_{t,n}(\epsilon) = \delta_n(\epsilon)$  when  $t(0) = 0.5$ , the capacity achieving input distribution. Therefore,

$$\begin{aligned} \max_{t \in \mathcal{P}(c)} [I(t; P) - \delta_{t,n}(\epsilon)] &= \max_{t \in \mathcal{P}(C_{\text{BIMSC}} - O(\delta_n(\epsilon)))} [I(t; P) - \delta_{t,n}(\epsilon)] \\ &= C_{\text{BIMSC}} - \delta_n(\epsilon) + O(\delta_n^2(\epsilon)). \end{aligned} \quad (3.101)$$

Consequently, by observing that the high order term  $o(\delta_n(\epsilon))$  in Theorem 2 is also in the order of  $\delta_n^2(\epsilon)$ , the Taylor-type expansion of  $R_n(\epsilon)$  for BIMSC in Theorem 4 is shown to be the same as that in Theorem 2.

#### D. Comparison with Asymptotic Analysis and Implication

It is instructive to compare Theorem 4 with the second order asymptotic performance analysis as  $n$  goes to  $\infty$ .

*Asymptotic analysis with constant  $0 < \epsilon < 1$  and  $n \rightarrow \infty$ :* Fix  $0 < \epsilon < 1$ . It was shown in [3], [4], [5] that for a DIMC with a discrete output alphabet and  $C_{\text{DIMC}} > 0$ ,

$$R_n(\epsilon) = C_{\text{DIMC}} - \frac{\sigma_D(P)}{\sqrt{n}} Q^{-1}(\epsilon) + O\left(\frac{\ln n}{n}\right) \quad (3.102)$$

for sufficiently large  $n$ , where

$$\sigma_D(P) = \begin{cases} \min\{\sigma_D(t; P) : t \in \mathcal{P} \& I(t; P) = C_{\text{DIMC}}\} & \text{if } \epsilon < \frac{1}{2} \\ \max\{\sigma_D(t; P) : t \in \mathcal{P} \& I(t; P) = C_{\text{DIMC}}\} & \text{if } \epsilon > \frac{1}{2}. \end{cases}$$

Once again, the expression  $C_{\text{DIMC}} - \frac{\sigma_D(P)}{\sqrt{n}} Q^{-1}(\epsilon)$  was referred to as the normal approximation for  $R_n(\epsilon)$  in [5]. It is not hard to verify that for sufficiently large  $n$ ,

$$\begin{aligned} C_{\text{DIMC}} - \frac{\sigma_D(P)}{\sqrt{n}} Q^{-1}(\epsilon) &\leq \max_{t \in \mathcal{P}(c)} \left[ I(t; P) - \frac{\sigma_D(t; P)}{\sqrt{n}} Q^{-1}(\epsilon) \right] \\ &= \max_{t: \exists p_X, |t - p_X| = O\left(\frac{1}{n^{1/2}}\right)} \left[ I(t; P) - \frac{\sigma_D(t; P)}{\sqrt{n}} Q^{-1}(\epsilon) \right] \\ &= C_{\text{DIMC}} - \frac{\sigma_D(P)}{\sqrt{n}} Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right) \end{aligned} \quad (3.103)$$

where the first equality is due to the fact that for any  $p_X$  satisfying  $I(p_X; P) = C_{\text{DIMC}}$  and  $t$  satisfying  $|t - p_X| = \omega(1/n^{1/2})$ ,

$$I(t; P) - \frac{\sigma_D(t; P)}{\sqrt{n}} Q^{-1}(\epsilon) \leq C_{\text{DIMC}} - \frac{\sigma_D(p_X; P)}{\sqrt{n}} Q^{-1}(\epsilon)$$

as

$$\frac{Q^{-1}(\epsilon)}{\sqrt{n}} |\sigma_D(t; P) - \sigma_D(p_X; P)| = O\left(\frac{|t - p_X|}{\sqrt{n}}\right) = o(|t - p_X|^2) = o(C_{\text{DIMC}} - I(t; P)).$$

Therefore, when  $\epsilon > 1/3$ , (3.102) and (3.84) are essentially the same for sufficiently large  $n$ .

Let us now look at the case  $\epsilon \leq 1/3$ . Again,  $0 < \epsilon \leq 1/3$  is fixed. In parallel with (2.109) and (2.110), we have for each  $t \in \mathcal{P}(c)$

$$r_-(t, \delta) = \frac{1}{2\sigma_D^2(t; p)} \delta^2 + \frac{-\hat{M}_D(t; P)}{6\sigma_D^6(t; P)} \delta^3 + O(\delta^4) \quad (3.104)$$

and

$$\delta_{t,n}(\epsilon) = \frac{\sigma_D(t; P)}{\sqrt{n}} Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right). \quad (3.105)$$

Combining (3.105) with (3.103) yields

$$\begin{aligned} C_{\text{DIMC}} - \frac{\sigma_D(P)}{\sqrt{n}} Q^{-1}(\epsilon) + O(1/n) &\leq \max_{t \in \mathcal{P}(c)} [I(t; P) - \delta_{t,n}(\epsilon)] \\ &\leq C_{\text{DIMC}} - \frac{\sigma_D(P)}{\sqrt{n}} Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right). \end{aligned} \quad (3.106)$$

Thus the Taylor-type expansion of  $R_n(\epsilon)$  in Theorem 4 implies the second order asymptotic analysis with constant  $0 < \epsilon < 1$  and  $n \rightarrow \infty$  shown in (3.102).

*Asymptotic analysis with  $n \rightarrow \infty$  and non-exponentially decaying  $\epsilon$ :* Suppose now  $\epsilon$  is a function of  $n$  and goes to 0 as  $n \rightarrow \infty$ , but at a non-exponential speed. Using arguments similar to those made above and in Subsection II-D, one can show that the Taylor-type expansion of  $R_n(\epsilon)$  in Theorem 4 implies that in this case,  $C_{\text{DIMC}}$  and  $-\frac{\sigma_D(P)}{\sqrt{n}} Q^{-1}(\epsilon)$  are still respectively the first order and second order terms of  $R_n(\epsilon)$  in the asymptotic analysis with  $n \rightarrow \infty$ . Once again, to the best of our knowledge, the second order asymptotic analysis with  $n \rightarrow \infty$  and non-exponentially decaying  $\epsilon$  has not been addressed before in the literature.

*Divergence from the normal approximation:* In the non-asymptotic regime where  $n$  is finite and  $\epsilon$  is generally relatively small with respect to  $n$ , the first two terms

$$\max_{t \in \mathcal{P}(c)} [I(t; P) - \delta_{t,n}(\epsilon)]$$

in the Taylor-type expansion of  $R_n(\epsilon)$  in Theorem 4 differ from the normal approximation in a strong way. In particular, the optimal distribution  $t^*$  defined in (3.99) is not necessarily a capacity achieving distribution. In this case, the normal approximation would fail to provide a reasonable estimate for  $R_n(\epsilon)$ .

*Example:* Consider the Z channel shown in Figure 1. In this example, we show that the

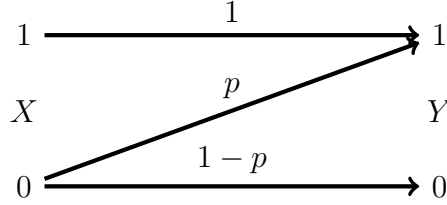


Fig. 1. Z Channel

optimal distribution  $t^*$  defined in (3.99) is not a capacity achieving distribution. In the numerical calculation shown in Figure 2, the transition probability  $p$  (i.e.  $\Pr\{Y = 1|X = 0\}$ ) ranges from 0.05 to 0.95 with block length  $n = 1000$  and error probability  $\epsilon = 10^{-6}$ . As can be seen from Figure 2(a),  $t^*(0)$  is always different from the capacity achieving  $t(0)$ . Moreover, Figure 2(b) shows the percentage of  $I(t; P) - \delta_{t,n}(\epsilon)$  over  $I(t^*; P) - \delta_{t^*,n}(\epsilon)$  when  $t$  is capacity achieving,  $t^*$ , and uniform respectively. It is clear that  $C_{\text{DIMC}} - \delta_{p_X,n}(\epsilon)$  is apart from  $I(t^*; P) - \delta_{t^*,n}(\epsilon)$  further and further when  $p$  gets larger and larger, where  $p_X$  is the capacity achieving distribution, indicating that under the practical block length and error probability requirement, Shannon random coding based on the capacity achieving distribution is not optimal. It is also interesting to note that for uniform  $t$ ,  $I(t; P) - \delta_{t,n}(\epsilon)$  is quite close to  $I(t^*; P) - \delta_{t^*,n}(\epsilon)$  within the whole range, implying that linear block coding is quite suitable for the Z channel even under the practical block length and error probability requirement.

*Implication on code design:* An important implication arising from the Taylor-type expansion of  $R_n(\epsilon)$  in Theorem 4 in the non-asymptotic regime is that for values of  $n$  and  $\epsilon$  with practical interest, the optimal marginal codeword symbol distribution is not necessarily a capacity achieving distribution. This is illustrated above for the Z channel. Indeed, other than for symmetric channels like BIMSC, it would expect that the optimal distribution  $t^*$  defined in (3.99) is in general not a capacity achieving distribution for values of  $n$  and  $\epsilon$  for which  $\delta_{t^*,n}(\epsilon)$  is not

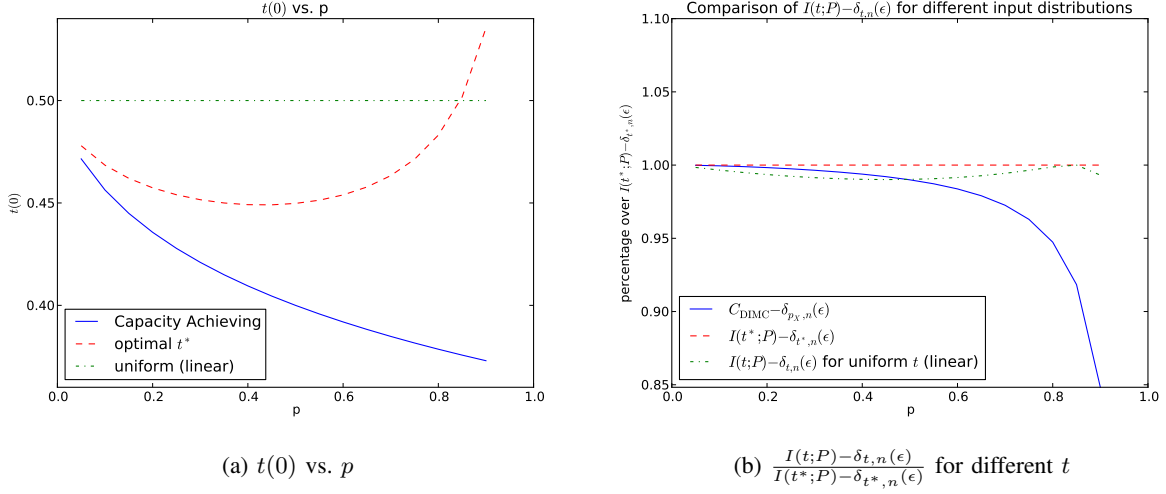


Fig. 2. Illustration for the Z channel with  $n = 1000$  and  $\epsilon = 10^{-6}$ : (a) comparison of  $t^*$  with the capacity achieving distribution; and (b) comparison of  $I(t;P) - \delta_{t,n}(\epsilon)$  among different distributions  $t$ .

relatively small. As such, to design efficient channel codes under the practical block length and error probability requirement, one approach is to solve the maximization problem in (3.99), get  $t^*$ , and then design codes so that the marginal codeword symbol distribution is approximately  $t^*$ .

#### IV. APPROXIMATION AND EVALUATION

Based on our converse theorems and Taylor-type expansion of  $R_n(\epsilon)$ , in this section, we first derive two approximation formulas for  $R_n(\epsilon)$ . We then compare them numerically with the normal approximation and some tight (achievable and converse) non-asymptotic bounds, for the BSC, BEC, BIAGC, and Z Channel. In all Figures 3 to 11, rates are expressed in bits.

##### A. Approximation Formulas

In view of the Taylor-type expansion of  $R_n(\epsilon)$  in Theorem 4, one reasonable approximation formula is to use the first two terms in Taylor-type expansion of  $R_n(\epsilon)$  as an estimate for  $R_n(\epsilon)$ . We refer to this formula as the second order (SO) formula:

$$\begin{aligned}
 R_n^{\text{SO}}(\epsilon) &= \max_{t \in \mathcal{P}(\mathcal{C})} [I(t;P) - \delta_{t,n}(\epsilon)] \\
 &= I(t^*;P) - \delta_{t^*,P}(\epsilon)
 \end{aligned} \tag{4.1}$$

where  $c$  is selected according to Remark 13.

To derive the other approximation formula for  $R_n(\epsilon)$ , let us put Theorem 3, Theorem 4, and the achievability given in (3.61) and (3.62) together. It would make sense for an optimal code of block length  $n$  to draw all its codewords from the same type  $t$  with  $|t - t^*| = O(1/n)$ . In this case, it is not hard to see that the term  $|\mathcal{X}|^{\frac{\ln(n+1)}{n}}$  in the bounds of Theorems 3 and 4 (i.e. (3.33), (3.35), (3.83), and (3.84)) can be dropped. By ignoring the higher order term  $\frac{\ln \frac{-2 \ln \epsilon n}{n} - \ln(1 + \sqrt{\frac{-2 \ln \epsilon n}{n}})}{n}$  in (3.33) and (3.35), we get the following approximation formula (dubbed “NEP”) :

$$R_n^{\text{NEP}}(\epsilon) = I(t^*; P) - \delta_{t^*; P}(\epsilon) - \frac{\ln \epsilon}{n} + \frac{1}{n} \ln P(B_{t^*, n, \delta_{t^*; P}(\epsilon)}) \quad (4.2)$$

Rewrite the normal approximation as

$$R_n^{\text{Normal}}(\epsilon) = C_{\text{DIMC}} - \frac{\sigma_D(P)}{\sqrt{n}} Q^{-1}(\epsilon). \quad (4.3)$$

## B. BIMSC

In the case of BIMSC, it follows from Theorem 2 and Remark 14 that  $R_n^{\text{SO}}(\epsilon)$ ,  $R_n^{\text{NEP}}(\epsilon)$ , and  $R_n^{\text{Normal}}(\epsilon)$  become respectively

$$\begin{aligned} R_n^{\text{SO}}(\epsilon) &= C_{\text{BIMSC}} - \delta_n(\epsilon) \\ R_n^{\text{NEP}}(\epsilon) &= C_{\text{BIMSC}} - \delta_n(\epsilon) - \frac{\ln \epsilon}{n} + \frac{1}{n} \ln P(B_{n, \delta_n(\epsilon)}) \end{aligned} \quad (4.4)$$

and

$$R_n^{\text{Normal}}(\epsilon) = C_{\text{BIMSC}} - \frac{\sigma_H(X|Y)}{\sqrt{n}} Q^{-1}(\epsilon). \quad (4.5)$$

From Theorem 2 and its comparison with asymptotic analysis, we can expect that when  $\delta_n(\epsilon)$  is extremely small,  $R_n^{\text{SO}}(\epsilon)$  and  $R_n^{\text{Normal}}(\epsilon)$  are close, and both can provide a good approximation for  $R_n(\epsilon)$ . However, as  $\delta_n(\epsilon)$  increases, the relative position of  $R_n^{\text{SO}}(\epsilon)$  and  $R_n^{\text{Normal}}(\epsilon)$  depends on

$$\zeta_{X|Y} = -\frac{\hat{M}_H(X|Y)}{6\sigma_H^6(X|Y)}.$$

Specifically, given a channel with large magnitude of  $\zeta_{X|Y}$ ,  $R_n^{\text{Normal}}(\epsilon)$  is not reliable, as it can be much below achievable bounds or above converse bounds. On the other hand, as shown later on,  $R_n^{\text{SO}}(\epsilon)$  is much more reliable. Moreover,  $R_n^{\text{NEP}}(\epsilon)$ , which has some terms beyond second order on top of  $R_n^{\text{SO}}(\epsilon)$ , always provides a good approximation for  $R_n(\epsilon)$  even if  $\delta_n(\epsilon)$  is relatively large.

1) *BSC*: For this channel, the trivial bound  $P(B_{n,\delta_n(\epsilon)}) \leq 1$  is applied in the evaluation of  $R_n^{\text{NEP}}(\epsilon)$ . Before jumping into the comparison of those approximations, let us first get some insight by investigating  $\zeta_{X|Y}$ . It can be easily verified that for BSC with cross-over probability  $p$ ,

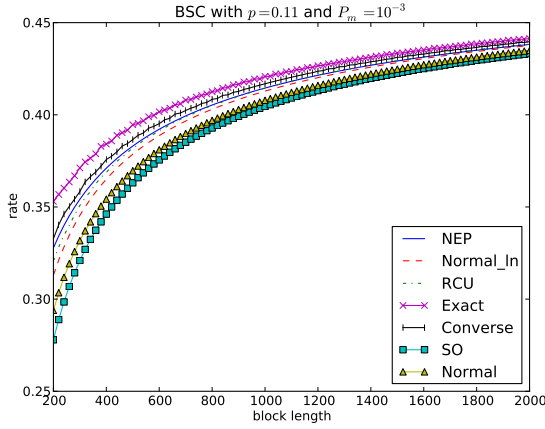
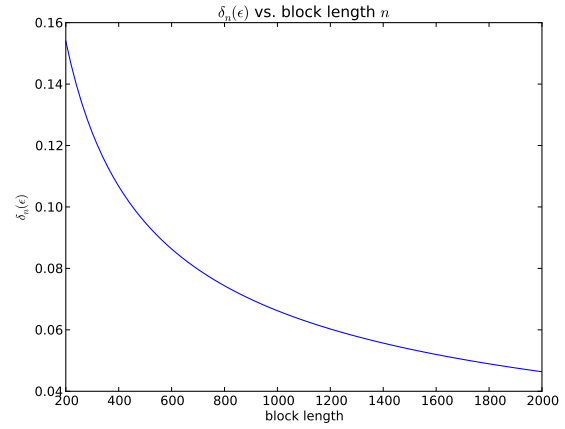
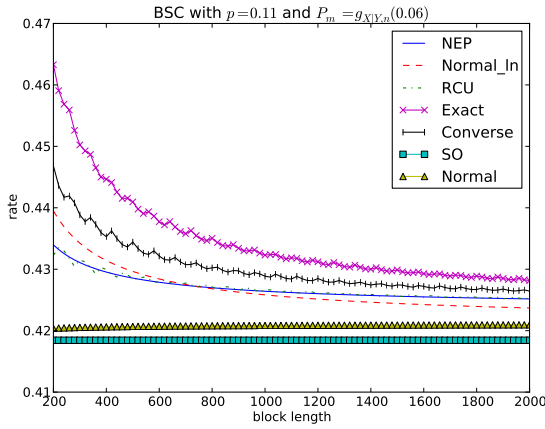
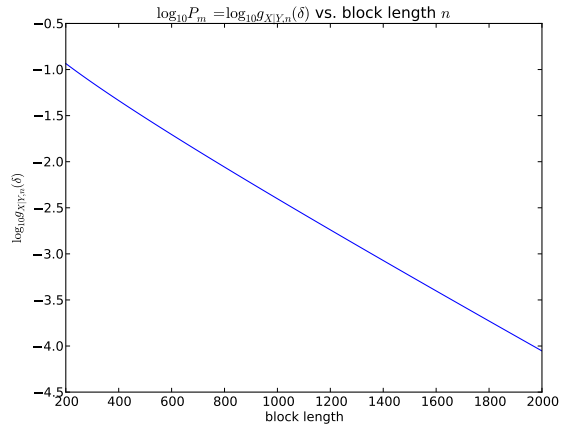
$$\zeta_{X|Y} = -\frac{1}{6 \ln^5 \frac{1-p}{p}} \frac{1-2p}{p^3(1-p)^3}. \quad (4.6)$$

As can be seen,  $\zeta_{X|Y}$  is always negative for any  $p \in (0, 1)$  and  $\zeta_{X|Y} \rightarrow -\infty$  as  $p \rightarrow 0$ . Therefore, in the case of a very small  $p$ ,  $R_n^{\text{Normal}}(\epsilon)$  will be larger than  $R_n^{\text{SO}}(\epsilon)$  by a relatively large margin, and even larger than the converse bound.

Now in order to compare those approximations, we invoke Theorem 33 (dubbed “RCU”) and Theorem 35 (dubbed “Converse”) in [5], which serve as an achievable bound and a converse bound, respectively. In addition, another converse bound is provided by the exact calculation of (2.60) and (2.61) in Corollary 1 (dubbed “Exact”). Moreover, by Theorem 52 in [5],  $\frac{\ln n}{2n}$  is the third order in the asymptotic analysis of  $R_n(\epsilon)$  as  $n \rightarrow \infty$  for BSC, and therefore, another approximation is yielded by adding  $\frac{\ln n}{2n}$  to the normal approximation (dubbed “Normal\_In”). Then these four approximation formulas (NEP, Normal\_In, Normal, SO), two converse bounds (Converse, Exact), and one achievable bound (RCU) are compared against each other with block length  $n$  ranging from 200 to 2000; their respective performance is shown in Figures 3 and 4.

In Figure 3, the target channel is the BSC with cross-over probability 0.11, where  $\zeta_{X|Y}$  is relatively small. In Figure 3(a), bounds are compared with fixed maximum error probability  $P_m = 10^{-3}$ , while  $\delta_n(\epsilon)$  changes with respect to block length  $n$ , shown in Figure 3(b). In the meantime, Figure 3(c) shows comparison of these bounds when  $\delta_n(\epsilon)$  is fixed to be 0.06, while  $P_m = g_{X|Y,n}(0.06)$  is shown in Figure 3(d). As can be seen, when  $\delta_n(\epsilon)$  gets smaller, the SO and Normal curves tend to coincide with each other. Moreover, since the SO and Normal approximation formulas are quite close in this case, both the NEP and Normal\_In provide quite accurate approximations for  $R_n(\epsilon)$  with the NEP slightly better.

Figure 4 shows the same curves as those in Figure 3, but for the BSC with cross-over probability 0.001. In this case, the magnitude of  $\zeta_{X|Y}$  is large, and therefore, the SO and Normal curves are well apart. In fact, the Normal curve is even above those two converse bounds, and so does the Normal\_In curve, thus confirming our analysis based on  $\zeta_{X|Y}$  made at the beginning of this discussion for BSC. On the other hand, the SO curve stays at the same relative position

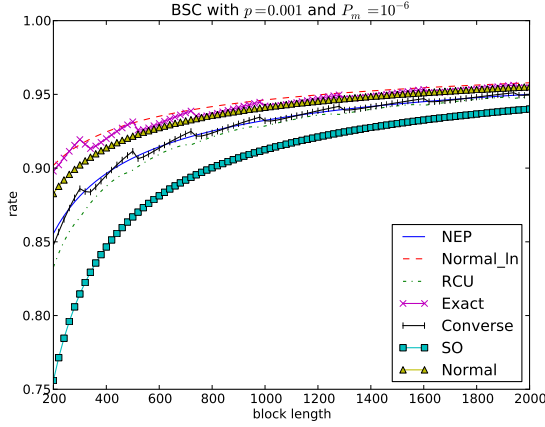
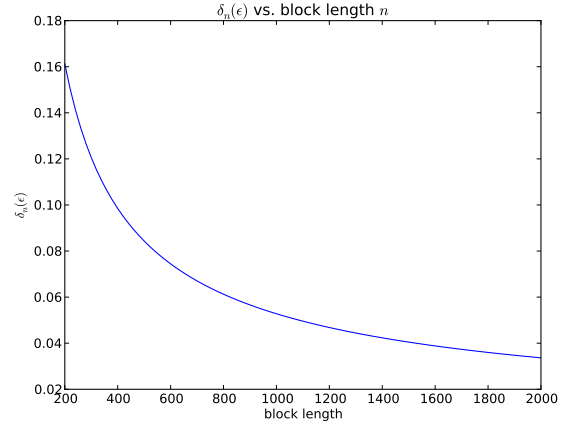
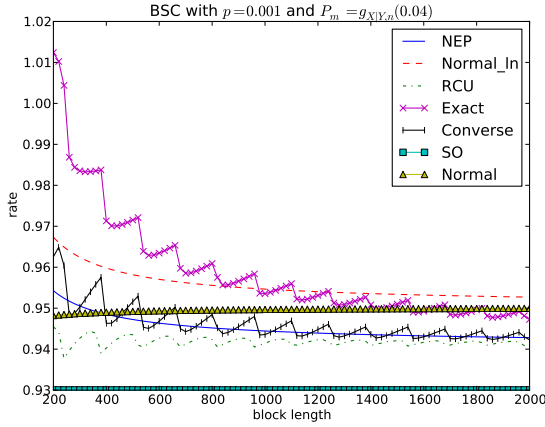
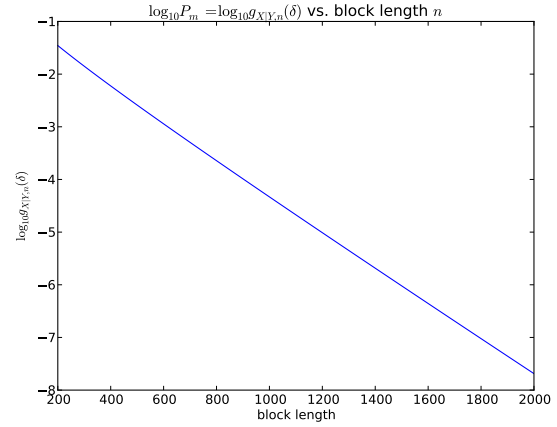
(a) Bounds with  $P_m = 10^{-3}$ (b)  $\delta_n(\epsilon)$  with  $P_m = 10^{-3}$ (c) Bounds with  $P_m = g_{X|Y,n}(\delta)$  and  $\delta = 0.06$ (d)  $\log_{10} P_m$  with  $P_m = g_{X|Y,n}(\delta)$  and  $\delta = 0.06$ Fig. 3. Comparison of different bounds for BSC with  $p = 0.11$ .

to achievable and converse bounds, and the NEP still provides an accurate approximation for  $R_n(\epsilon)$ .

2) *BEC*: This special channel serves as another interesting example to illustrate the difference between the SO and Normal approximations. On one hand, it can be easily verified that

$$P(B_{n,\delta}) = \Pr \left\{ -\frac{1}{n} \ln p(X^n|Y^n) > H(X|Y) + \delta \right\} \approx g_{X|Y,n}(\delta) \quad (4.7)$$



(a) Bounds with  $P_m = 10^{-6}$ (b)  $\delta_n(\epsilon)$  with  $P_m = 10^{-6}$ (c) Bounds with  $P_m = g_{X|Y,n}(\delta)$  and  $\delta = 0.04$ (d)  $\log_{10} P_m$  with  $P_m = g_{X|Y,n}(\delta)$  and  $\delta = 0.04$ Fig. 4. Comparison of different bounds for BSC with  $p = 0.001$ .

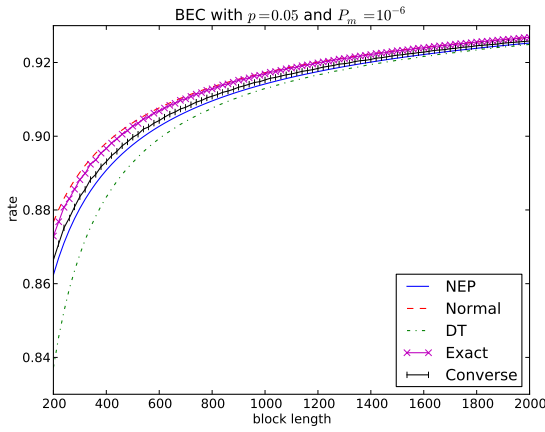
and therefore,  $-\frac{\ln \epsilon}{n}$  and  $\frac{1}{n} \ln P(B_{n,\delta_n(\epsilon)})$  are cancelled out in  $R_n^{\text{NEP}}(\epsilon)$ , which is then identical to  $R_n^{\text{SO}}(\epsilon)$ . On the other hand,

$$\zeta_{X|Y} = -\frac{(1-2p)}{6p^2(1-p)^2 \ln^3 2} \begin{cases} < 0 & \text{if } p < 0.5 \\ = 0 & \text{if } p = 0.5 \\ > 0 & \text{if } p > 0.5 \end{cases} \quad (4.8)$$

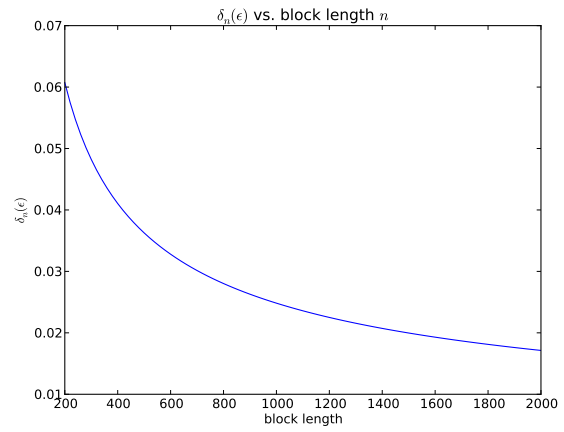
Therefore, the Normal curve can be all over the map, i.e. it can be above some converse when  $p < 0.5$ , and below an achievable bound when  $p > 0.5$ . When  $p = 0.5$ , the Normal curve happens to be close to the SO curve, hereby explaining why it provides an accurate approximation for

$R_n(\epsilon)$  in this particular case, as shown in [5].

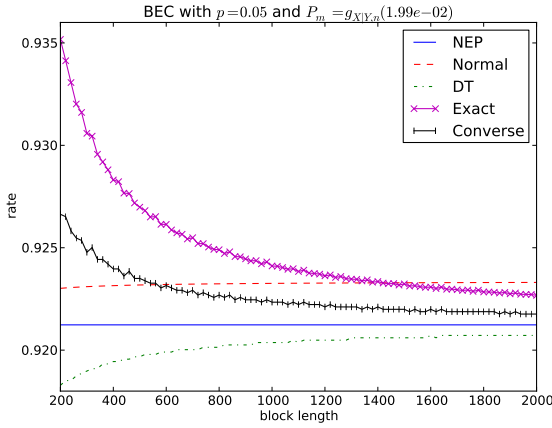
To provide benchmarks for the comparison of approximation formulas, Theorem 37 and 38 in [5] are used here, dubbed “DT” and “Converse” respectively. The exact calculation of (2.60) and (2.61) in Corollary 1 (dubbed “Exact”) again serves as an additional converse bound. Then those bounds are drawn in Figures 5 and 6 in the same way as those in figure 3, where erasure probabilities are selected to be 0.05 and 0.9, respectively. Once again, numeric results confirm our analysis and discussion above.



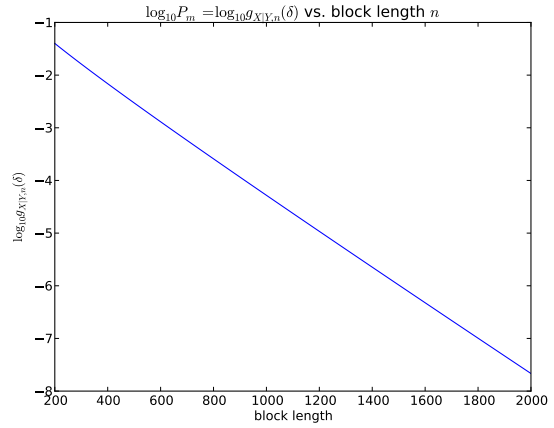
(a) Bounds with  $P_m = 10^{-6}$



(b)  $\delta_n(\epsilon)$  with  $P_m = 10^{-6}$



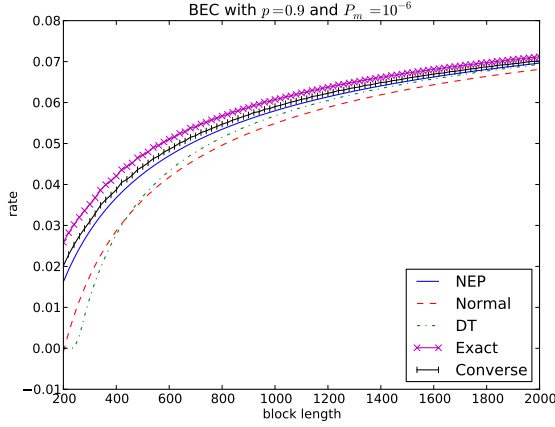
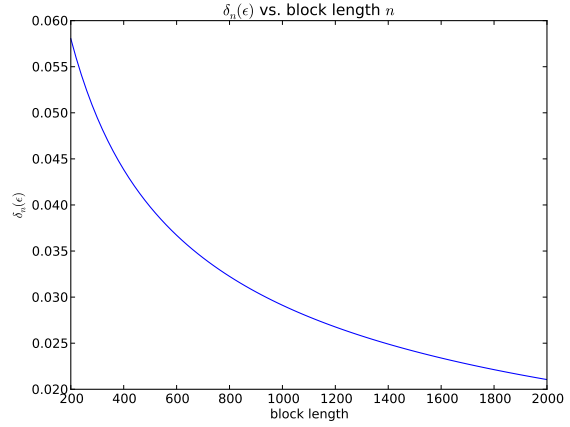
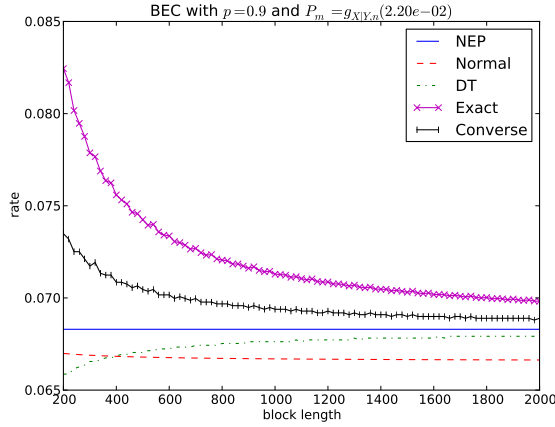
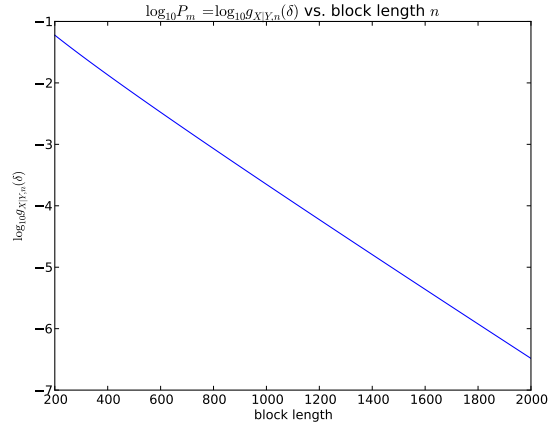
(c) Bounds with  $P_m = g_{X|Y,n}(\delta)$  and  $\delta = 0.0199$



(d)  $\log_{10} P_m$  with  $P_m = g_{X|Y,n}(\delta)$  and  $\delta = 0.0199$

Fig. 5. Comparison of different bounds for BEC with  $p = 0.05$ .

3) *BIAGC*: Here we assume that codewords are modulated to  $\{+1, -1\}$  before going through an AWGN channel, and apply the trivial bound  $P(B_{n,\delta_n(\epsilon)}) \leq 1$  in the NEP formula. Similarly

(a) Bounds with  $P_m = 10^{-6}$ (b)  $\delta_n(\epsilon)$  with  $P_m = 10^{-6}$ (c) Bounds with  $P_m = g_{X|Y,n}(\delta)$  and  $\delta = 0.022$ (d)  $\log_{10} P_m$  with  $P_m = g_{X|Y,n}(\delta)$  and  $\delta = 0.022$ Fig. 6. Comparison of different bounds for BEC with  $p = 0.9$ .

to BSC and BEC, we would like to get some insight by investigating  $\zeta_{X|Y}$ . Since in this case,  $\zeta_{X|Y}$  does not seem to have a simple close form expression which can be easily computed, numerical calculation of  $\zeta_{X|Y}$  is shown in Figure 7, where SNR ranges from 8dB to 10.5dB. As can be seen, BIAGC is similar to BSC, i.e.  $\zeta_{X|Y}$  is always negative and its magnitude increases with SNR. Therefore,  $R_n^{\text{Normal}}(\epsilon)$  is close to  $R_n^{\text{SO}}(\epsilon)$  when SNR is low, but can be above some converse bounds when SNR is high. This is confirmed in Figures 8 and 9, where exact evaluation of (2.62) and (2.63) in Corollary 1 (dubbed “Exact”) serves as a converse bound.

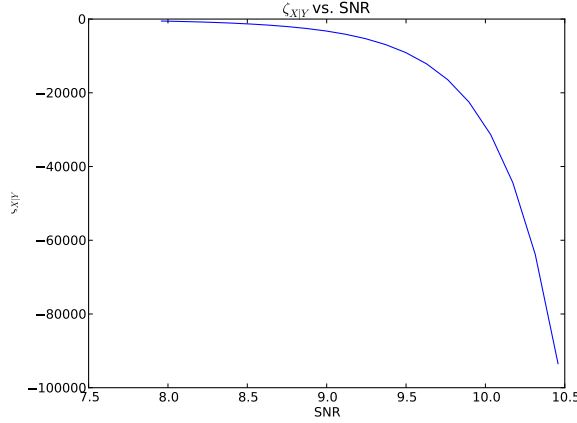


Fig. 7.  $\zeta_{X|Y}$  of BIAGC

### C. DIMC: Z Channel

To show an example of DIMC which is not a BIMSC, we consider again the Z channel shown in Figure 1. The capacity of Z channel is well known and given by

$$C_Z = \ln \left( 1 + (1-p)p^{\frac{p}{1-p}} \right) \quad (4.9)$$

with the capacity-achieving distribution

$$p_X(x) = \begin{cases} \frac{1}{1-p+p^{-\frac{p}{1-p}}} & \text{for } x = 0 \\ \frac{p^{-\frac{p}{1-p}}-p}{1-p+p^{-\frac{p}{1-p}}} & \text{for } x = 1 \end{cases} \quad (4.10)$$

and the corresponding output distribution

$$p_Y(y) = \begin{cases} \frac{1-p}{1-p+p^{-\frac{p}{1-p}}} & \text{for } y = 0 \\ \frac{p^{-\frac{p}{1-p}}}{1-p+p^{-\frac{p}{1-p}}} & \text{for } y = 1. \end{cases} \quad (4.11)$$

To calculate  $R_n^{\text{NEP}}(\epsilon)$ ,  $P(B_{t,n,\delta})$  needs to be further investigated, where an interesting observation is that given  $x^n$  with type  $t$ ,  $\frac{1}{n} \ln \frac{p(y^n|x^n)}{q_t(y^n)} > -\infty$  if and only if  $y_i = 1$  when  $x_i = 1$ , and the value of  $\frac{1}{n} \ln \frac{p(y^n|x^n)}{q_t(y^n)}$  only depends on the number of  $y_i$  being 1 for  $i \in \{j : x_j = 0\}$ . One can then verify that

$$B_{t,n,\delta} = \left\{ y^n : \frac{1}{n} |\{i : y_i = 0\}| \leq q_t(0) - \frac{\delta}{\ln \frac{1-t(0)+pt(0)}{pt(0)}} \right\}. \quad (4.12)$$

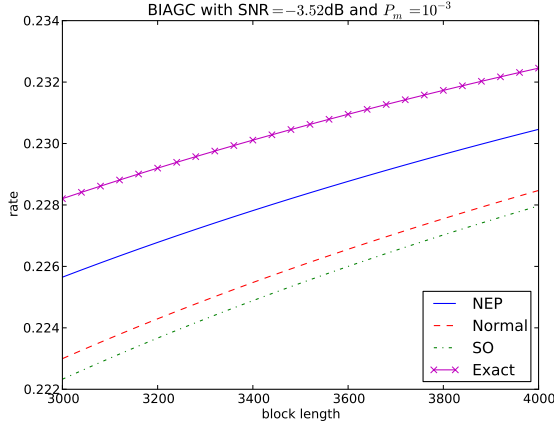
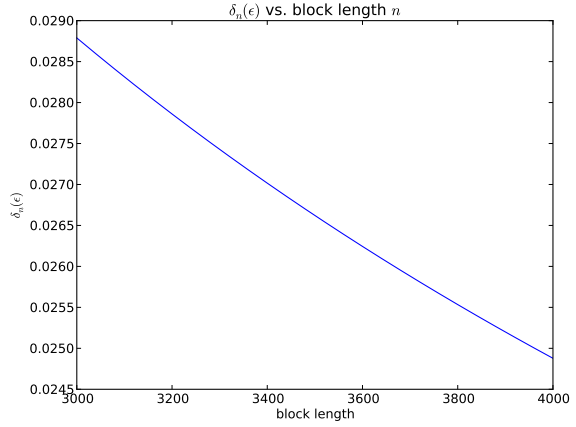
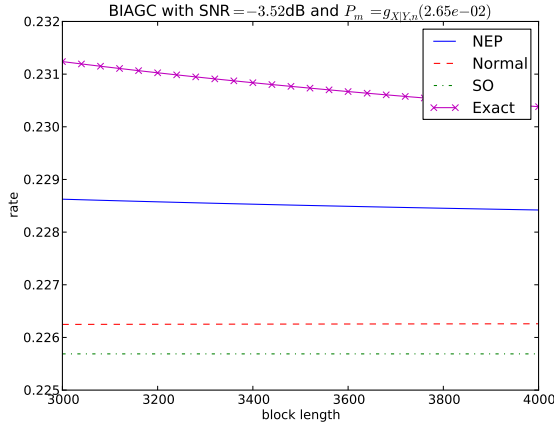
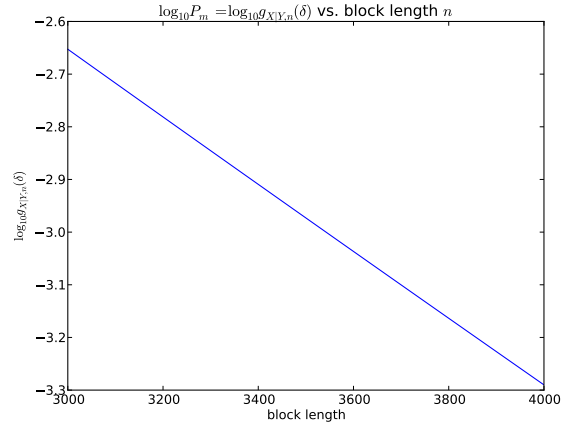
(a) Bounds with  $P_m = 10^{-3}$ (b)  $\delta_n(\epsilon)$  with  $P_m = 10^{-3}$ (c) Bounds with  $P_m = g_{X|Y,n}(\delta)$  and  $\delta = 0.0265$ (d)  $\log_{10} P_m$  with  $P_m = g_{X|Y,n}(\delta)$  and  $\delta = 0.0265$ 

Fig. 8. Comparison of different bounds for BIAGC with SNR = -3.52 dB.

When  $q_t(0) \neq 0.5$ ,

$$P(B_{t,n,\delta}) = \begin{cases} \Pr \left\{ -\frac{1}{n} \ln q_t(Y_t^n) \leq H(Y_t) - \frac{\delta}{\ln \frac{1-t(0)+pt(0)}{pt(0)}} \ln \frac{1-q_t(0)}{q_t(0)} \right\} & \text{if } q_t(0) < 0.5 \\ \Pr \left\{ -\frac{1}{n} \ln q_t(Y_t^n) \geq H(Y_t) - \frac{\delta}{\ln \frac{1-t(0)+pt(0)}{pt(0)}} \ln \frac{1-q_t(0)}{q_t(0)} \right\} & \text{if } q_t(0) > 0.5 \end{cases} \quad (4.13)$$

where  $Y_t$  is a random variable with distribution  $q_t$ . Consequently, we can apply the left NEP [6], chernoff bound, right NEP [6] with respect to entropy to upper bound  $P(B_{t,n,\delta})$  when  $q_t(0) <, =, > 0.5$ , respectively.

To provide benchmarks for the comparison of approximation formulas, exact evaluation of (3.55) (with  $|\mathcal{X}| \frac{\ln(n+1)}{n}$  dropped and  $t = t^*$ ) and (3.56) is provided, which, dubbed “Exact”,

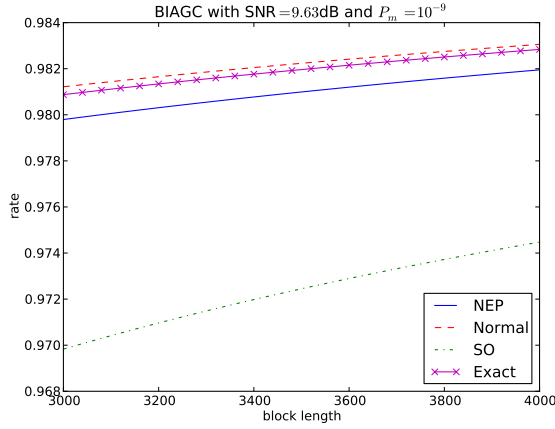
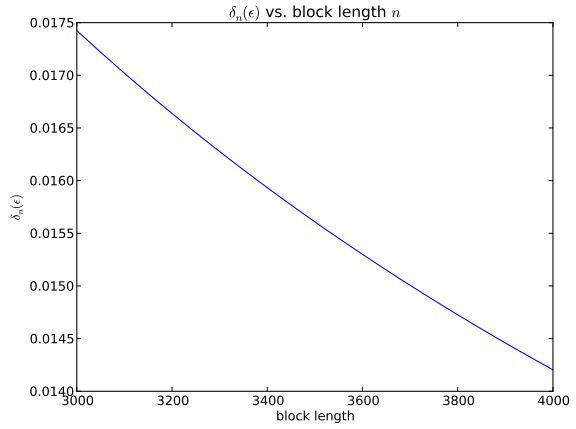
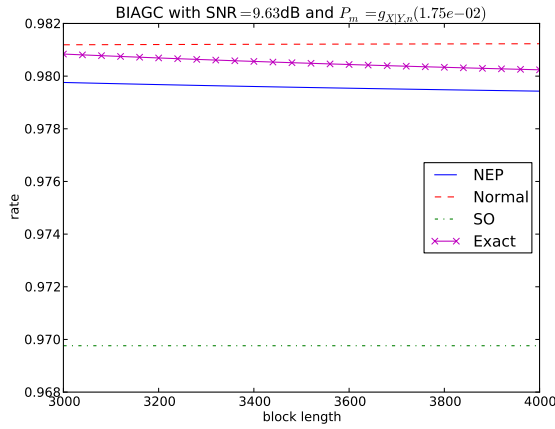
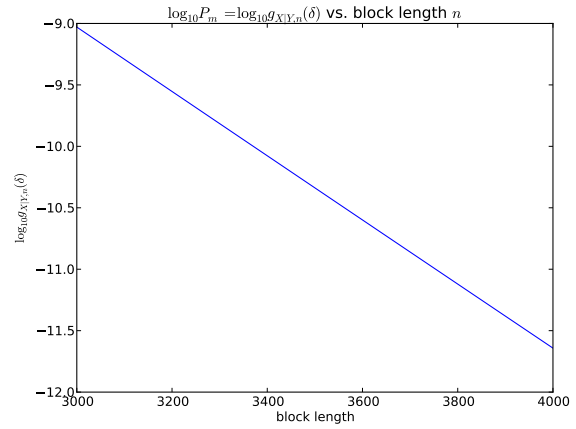
(a) Bounds with  $P_m = 10^{-9}$ (b)  $\delta_n(\epsilon)$  with  $P_m = 10^{-9}$ (c) Bounds with  $P_m = g_{X|Y,n}(\delta)$  and  $\delta = 0.0175$ (d)  $\log_{10} P_m$  with  $P_m = g_{X|Y,n}(\delta)$  and  $\delta = 0.0175$ 

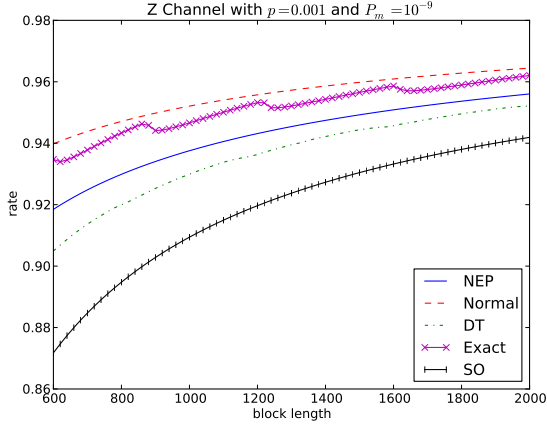
Fig. 9. Comparison of different bounds for BIAGC with SNR = 9.63 dB.

serves as a converse bound, and Theorem 22 in [5] provides an achievable bound, dubbed “DT” and given below:

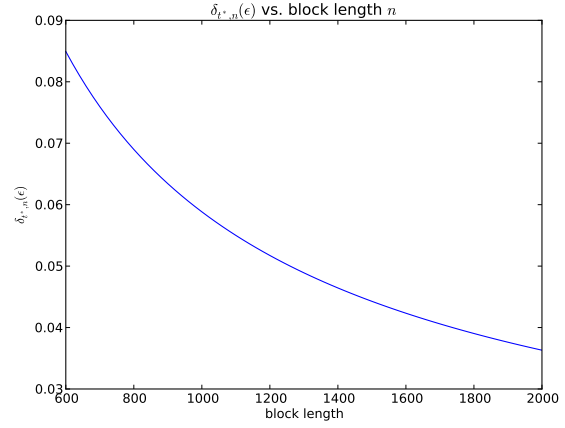
$$P_m \leq \sum_{i=0}^m \binom{m}{i} (1-p)^{m-i} p^i \min \left\{ 1, (M-1) \frac{\binom{n-m+i}{i}}{\binom{n}{m}} \right\} \quad (4.14)$$

where  $M = 2^{nR}$  and  $m = t^*(0)n$ . Figures 10 and 11 again show that the Normal curve is all over the map while the NEP curve always lies in between the DT achievable curve and the Exact

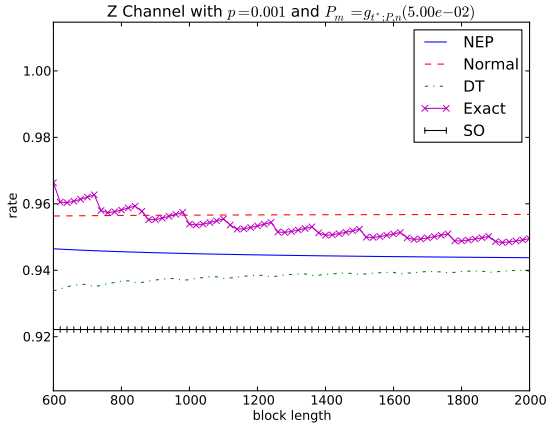
converse curve. It is also worth pointing out that if the capacity achieving distribution  $t = p_X$  instead of  $t^*$  was chosen in the calculation of the Exact and DT bounds, then both of them would be lower, confirming our early discussion that in the practical, non-asymptotic regime, the optimal marginal codeword symbol distribution is not necessarily a capacity achieving distribution.



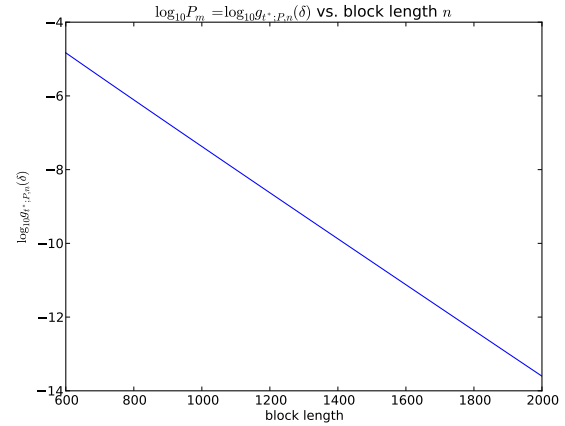
(a) Bounds with  $P_m = 10^{-9}$



(b)  $\delta_{t^*,n}(\epsilon)$  with  $P_m = n^{-9}$



(c) Bounds with  $P_m = g_{t^*,P,n}(\delta)$  and  $\delta = 0.05$

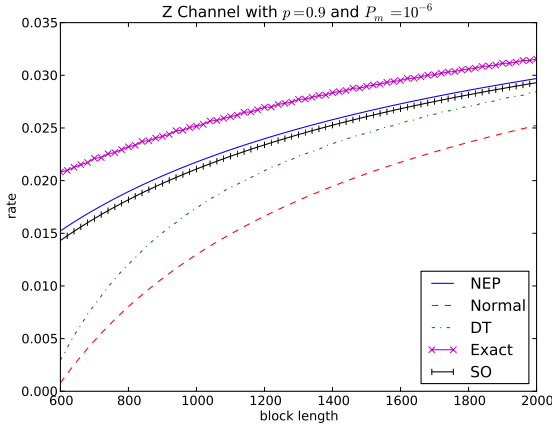
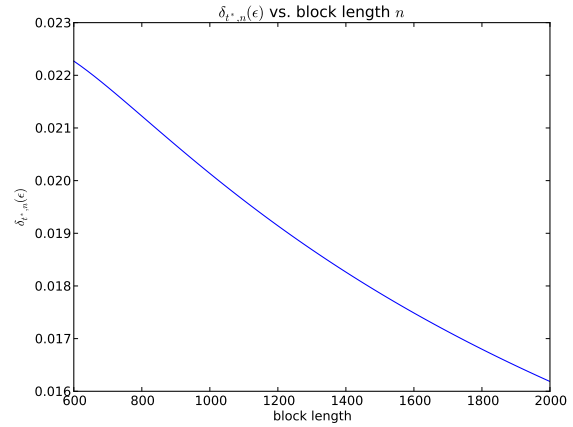
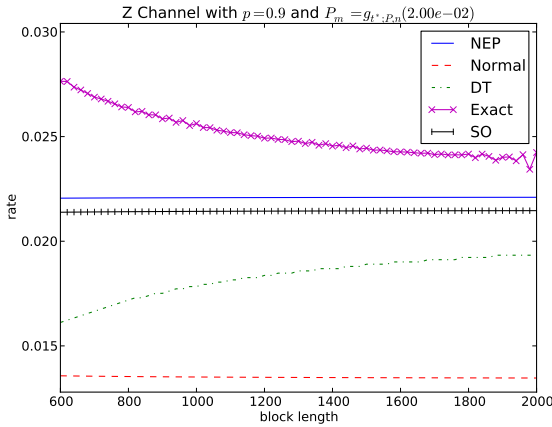
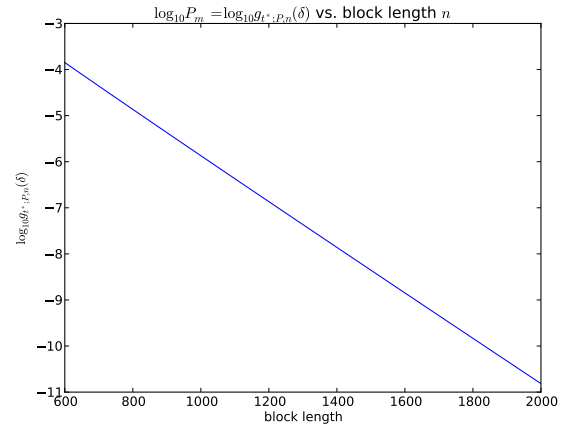


(d)  $\log_{10} P_m$  with  $P_m = g_{t^*,P,n}(\delta)$  and  $\delta = 0.05$

Fig. 10. Comparison of different bounds for Z Channel with  $p = 0.001$ .

## V. CONCLUSION

In this paper, we have developed a new converse proof technique dubbed the outer mirror image of jar and used it to establish new non-asymptotic converses for any discrete input memoryless channel with discrete or continuous output. Combining these non-asymptotic converses with the

(a) Bounds with  $P_m = 10^{-6}$ (b)  $\delta_{t^*,n}(\epsilon)$  with  $P_m = n^{-6}$ (c) Bounds with  $P_m = g_{t^*,P,n}(\delta)$  and  $\delta = 0.02$ (d)  $\log_{10} P_m$  with  $P_m = g_{t^*,P,n}(\delta)$  and  $\delta = 0.02$ Fig. 11. Comparison of different bounds for Z Channel with  $p = 0.9$ .

non-asymptotic achievability proved in [1] and [2] under jar decoding and with the NEP technique developed recently in [6], we have characterized the best coding rate  $R_n(\epsilon)$  achievable with finite block length  $n$  and error probability  $\epsilon$  through introducing a quantity  $\delta_{t,n}(\epsilon)$  to measure the relative magnitude of the error probability  $\epsilon$  and block length  $n$  with respect to a given channel  $P$  and an input distribution  $t$ . We have showed that in the non-asymptotic regime where both  $n$  and  $\epsilon$  are finite,  $R_n(\epsilon)$  has a Taylor-type expansion with respect to  $\delta_{t,n}(\epsilon)$ , where the first two terms of the expansion are  $\max_t [I(t; P) - \delta_{t,n}(\epsilon)]$ , which is equal to  $I(t^*, P) - \delta_{t^*,n}(\epsilon)$  for some optimal distribution  $t^*$ , and the third order term of the expansion is  $O(\delta_{t^*,n}^2(\epsilon))$  whenever  $\delta_{t^*,n}(\epsilon) =$



$\Omega(\sqrt{\ln n/n})$ . Based on the new non-asymptotic converses and the Taylor-type expansion of  $R_n(\epsilon)$ , we have also derived two approximation formulas (dubbed “SO” and “NEP”) for  $R_n(\epsilon)$ . These formulas have been further evaluated and compared against some of the best bounds known so far, as well as the normal approximation revisited recently in the literature. It turns out that while the normal approximation is all over the map, i.e. sometime below achievability and sometime above converse, the SO approximation is much more reliable and stays at the same relative position to achievable and converse bounds; in the meantime, the NEP approximation is the best among the three and always provides an accurate estimation for  $R_n(\epsilon)$ .

It is expected that in the non-asymptotic regime where both  $n$  and  $\epsilon$  are finite, the Taylor-type expansion of  $R_n(\epsilon)$  and the NEP approximation formula would play a role similar to that of Shannon capacity [8] in the asymptotic regime as  $n \rightarrow \infty$ . For values of  $n$  and  $\epsilon$  with practical interest for which  $\delta_{t^*,n}(\epsilon)$  is not relatively small, the optimal distribution  $t^*$  achieving  $\max_t [I(t; P) - \delta_{t,n}(\epsilon)]$  is in general not a capacity achieving distribution except for symmetric channels such as binary input memoryless symmetric channels. As a result, an important implication arising from the Taylor-type expansion of  $R_n(\epsilon)$  is that in the practical non-asymptotic regime, the optimal marginal codeword symbol distribution is not necessarily a capacity achieving distribution. Therefore, it will be interesting to examine all practical channel codes proposed so far against the Taylor-type expansion of  $R_n(\epsilon)$  and the NEP approximation formula and to see how far their performance is away from that predicted by the Taylor-type expansion of  $R_n(\epsilon)$  and the NEP approximation formula. If the performance gap is significant, one way to design a better channel code with practical block length and error probability requirement is to solve the maximization problem  $\max_t [I(t; P) - \delta_{t,n}(\epsilon)]$ , get  $t^*$ , and then design a code so that its marginal codeword symbol distribution is approximately  $t^*$ .

Finally, we conclude this paper by saying a few words on non-asymptotic information theory. From the viewpoint of stochastic processes, most classic results in information theory are based, to a large extent, on the strong and weak laws of large numbers and on large deviation theory. For example, most first order asymptotic coding rate results in information theory were established through the applications of asymptotic equipartition properties and typical sequences [9], which in turn depend on the strong and weak laws of large numbers. On other hand, error exponent analysis in both source and channel coding is in the spirit of large deviation theory. The recent second order asymptotic coding rate results [3], [4], [5] depend heavily on the Berry-Esseen

central limit theorem. In the non-asymptotic regime of practical interest, however, none of these probabilistic tools can be applied directly. To fill in this void space, we have developed the NEP in [6]. Based on the NEP, we have further invented jar decoding in [1] and presented the outer mirror image of jar converse proof technique in this paper. As demonstrated in this paper along with [1] and [6], the NEP, jar decoding, and the outer mirror image of jar together form a set of essential techniques needed for non-asymptotic information theory. They can also be extended and applied to help develop non-asymptotic multi-user information theory as well.

## REFERENCES

- [1] E.-H. Yang and J. Meng, “Jar decoding: Basic concepts and non-asymptotic capacity achieving coding theorems for channels with discrete inputs,” submitted to *IEEE Trans. on Inform. Theory*, 2011.
- [2] —, “Jar decoding: Basic concepts and non-asymptotic capacity achieving linear coding theorems,” submitted to *IEEE International Symposium on Information Theory*, 2012.
- [3] V. Strassen, “Asymptotische abschätzungen in shannon’s informationstheorie,” in *Proc. 3rd Conf. Inf. Theory*, Prague, Czech Republic, 1962, pp. 689–723.
- [4] M. Hayashi, “Information spectrum approach to second-order coding rate in channel coding,” *Information Theory, IEEE Transactions on*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.
- [5] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [6] E.-H. Yang and J. Meng, “Non-asymptotic equipartition properties for independent and identically distributed sources,” submitted to *IEEE Trans. on Inform. Theory*, 2011.
- [7] I. Csiszar, “The method of types,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, Oct. 1998.
- [8] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July, October 1948.
- [9] T.-M. Cover and J.-A. Thomas, *Elements of Information Theory (second edition)*. Hoboken, NJ: Wiley, 2006.